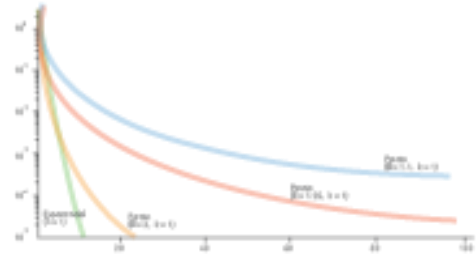


IMPLEMENTACIÓN DEL MODELO CASUAL PARA LA ESTIMACIÓN DE LA DEMANDA DE TRÁFICO DE INTERNET EN ALGUNOS PAÍSES DE LA REGIÓN SURAMERICANA

Jorge Iván Posada G.

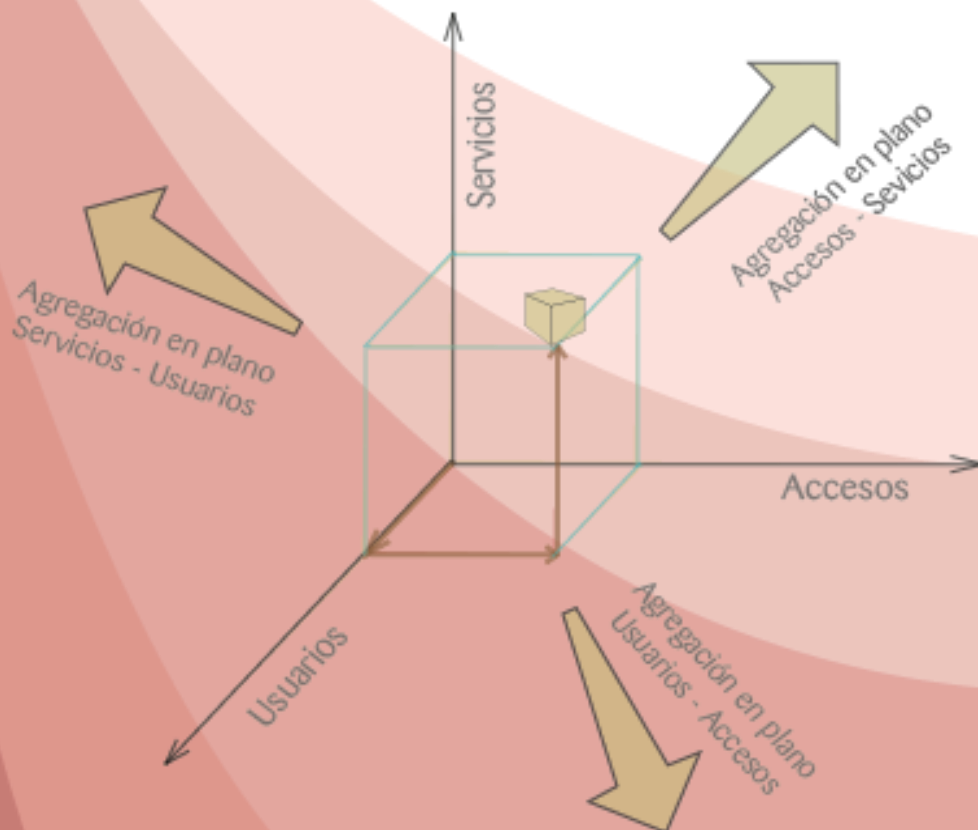
Used Internet Connections
By Country (2010)





IMPLEMENTACIÓN DEL MODELO CASUAL PARA LA ESTIMACIÓN DE LA DEMANDA DE TRÁFICO DE INTERNET EN ALGUNOS PAÍSES DE LA REGIÓN SURAMERICANA

Jorge Iván Posada G.



CALIFICACIONES



Natalia Gaviria Gómez

Asesora de proyecto

Erwin Alexander Leal

Asesor de proyecto

UNIVERSIDAD
DE ANTIOQUIA
1803

Ana María Cárdenas

Jurado designado

**IMPLEMENTACIÓN DEL MODELO CASUAL
PARA LA ESTIMACIÓN DE LA DEMANDA DE INTERNET
EN ALGUNOS PAÍSES DE LA REGIÓN SURAMERICANA**

TESIS DE GRADO PARA OPTAR POR EL TÍTULO

DE INGENIERO ELECTRÓNICO

JORGE IVÁN POSADA GÁMEZ

BAJO LA DIRECCIÓN DE

NATALIA GAVIRIA GÓMEZ

DOCTORA DE INGENIERÍA ELÉCTRICA Y DE COMPUTADORES

Y

ERWIN ALEXANDER LEAL

INGENIERO ELECTRÓNICO CON ENFASIS EN TELECOMUNICACIONES

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

Departamento de Ingeniería Electrónica

Medellín, Marzo 2012



Dedicado a todos los que han tendido su mano,
me han cargado en sus hombros,
y han creído en mi.

Éste es el fruto de su esfuerzo,
ahora viene el tiempo de la cosecha.

UNIVERSIDAD
DE ANTIOQUIA

1803

A decorative orange line art pattern at the bottom of the page, consisting of a series of vertical lines of varying heights and thicknesses, creating a textured, abstract effect.

AGRADECIMIENTOS

Agradezco a Dios la vida que me ha dado y que me ha concedido disfrutar, a mi madre por el amor con que guió a pesar de mis constantes rebeldías, a mi padre por su apoyo en esta etapa de mi vida, a mi tío Gildardo por sus innumerables auxilios, a mi tío Fernando por su permanente ayuda y finalmente a todos los que directa o indirectamente me han colaborado para alcanzar esta meta y cerrar este ciclo.

Agradezco a Ian Karlo Velez y al equipo de Planeación y Desarrollo de Negocios de Internexa S.A. E.S.P. por la oportunidad de trabajar con ellos y quienes con sus inquietantes preguntas me ayudaron y motivaron a dar los primeros pasos en la ejecución de este proyecto en mi periodo de práctica del año 2010.

¡Gracias!

UNIVERSIDAD
DE ANTIOQUIA
1803



ABSTRACT

This work aims at estimating the demand for Internet traffic in some of the countries of the South American region, Colombia, Ecuador, Peru, Chile, Bolivia, Uruguay, Paraguay, Argentina, Brazil and Venezuela, from the web service and some of its contents, images, text codes.

It review the classification and evolution of different existing models for calculating Teletraffic demand, we study the model On - Off, self-similar models, Poisson processes, fractional Brownian motion among others.

It interprets and implements the model fits **CASUAL** (*Cube of Access, Service, and User Allocation Free*), in the case of downloading and viewing Web pages. It explains the three axes on which the model is supported **CASUAL** (axle access, axle service, and axle users).

UNIVERSIDAD
DE ANTIOQUIA
1803



RESUMEN

Con este trabajo se pretende calcular la demanda de tráfico de Internet en algunos de los países de la región suramericana, Colombia, Ecuador, Perú, Chile, Bolivia, Uruguay, Paraguay, Argentina, Brasil y Venezuela, a partir del servicio web y algunos de sus contenidos, imágenes, texto, códigos.

Se revisa la clasificación y evolución de distintos modelos existentes para el cálculo de la demanda de Teletráfico, se estudia el modelo On - Off, modelos de autosimilitud, procesos de Poisson, movimiento fraccional browniano entre otros.

Se interpreta, adapta e implementa el modelo **CASUAL** (*Cubo de Acceso, Servicio y Usuario de Asignación Libre*), en el caso de descarga y visualización de páginas web. Se explican los tres ejes sobre los que se soporta el modelo **CASUAL** (ejes de accesos, eje de servicios y eje de usuarios).

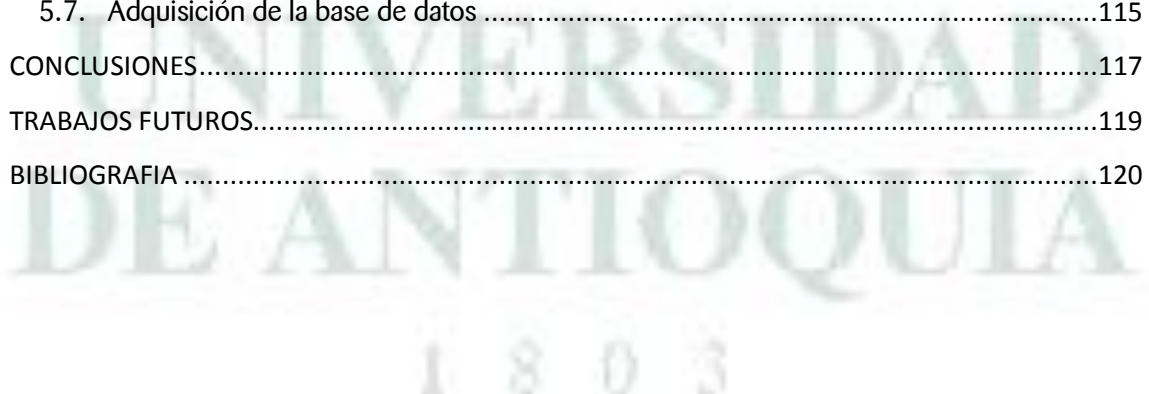
UNIVERSIDAD
DE ANTIOQUIA
1803



ÍNDICE DE CONTENIDO

CALIFICACIONES.....	3
ABSTRACT	7
RESUMEN	8
ÍNDICE DE CONTENIDO.....	9
ÍNDICE DE FIGURAS	11
ÍNDICE DE TABLAS	12
ÍNDICE DE ACRÓNIMOS.....	13
INTRODUCCIÓN	16
Convenciones tipográficas.....	18
OBJETIVOS	20
Objetivo general.....	20
Objetivos específicos.....	20
1. MODELOS DE TELETRÁFICO	21
1.1. Breve Historia	21
1.2. Tipos de modelos de Teletráfico	22
1.3. Algunos modelos usados en teletráfico.....	27
1.3.1 Modelos de Poisson.....	28
1.3.2 Modelos de Bernoulli.....	28
1.3.3 Modelo tipo fase.....	29
1.3.4 Modelo lineal autoregresivo.....	29
1.3.5 Modelo de Equilibrio de Wardrop	29
1.3.5 Procesos TES.....	30
1.3.6 Modelo movimiento fraccional browniano.....	30
1.3.7 Modelos On-Off	30
2. EJE DE SERVICIOS.....	35
2.1 Sobre la comunicación de datos y las redes	36
2.1.1 Arquitectura Cliente / Servidor	36
2.1.2 El modelo OSI y le modelo de Internet	48

2.1.3 Internet y TCP/IP.....	57
2.2 La WEB	63
2.2.1 ¿Qué es un sitio web?	64
2.2.2 ¿Qué es una aplicación web?	64
2.2.3 Tipos de objetos	65
3. EJE DE ACCESOS	70
3.1 Tecnologías de acceso fijas	70
3.2 Tecnologías de acceso inalámbricas.....	71
3.3 Los servicios de Internet	73
4. EJE DE USUARIOS	75
5. MODELO CASUAL	79
5.2 Ventajas y desventajas.....	80
5.3. Características del modelo CASUAL.....	81
5.3.1. Características del modelo ON-OFF en el modelo CASUAL.....	84
5.3.2. Distribuciones Estadísticas	86
5.4. Parámetros principales del modelo CASUAL.....	93
5.5. Interpretación de la formulación del modelo CASUAL.....	100
5.6. Resultados de la implementación del modelo	105
5.6.1. Parámetros y capacidad requerida de las páginas más relevantes	105
5.6.2. Parámetros y capacidad requerida para N fuentes	108
5.6.3. Capacidad equivalente del servicio web en distintos años de los países en estudio	109
5.6.4. Capacidad total a partir del servicio web	110
5.7. Adquisición de la base de datos	115
CONCLUSIONES.....	117
TRABAJOS FUTUROS.....	119
BIBLIOGRAFIA	120



ÍNDICE DE FIGURAS

<i>Figura 1. Representación del tráfico On-Off, (nivel de paquetes)</i>	31
<i>Figura 2. Traza On-Off de tráfico MMS</i>	32
<i>Figura 3. Representación de una sesión en el modelo On-Off</i>	32
<i>Figura 4. Arquitectura básica Cliente/Servidor</i>	36
<i>Figura 5. Arquitectura todo en un servidor</i>	39
<i>Figura 6. Un sólo servidor con servicio de aplicaciones</i>	39
<i>Figura 7. Servidor de base de datos por aparte</i>	40
<i>Figura 8. Sólo dos servidores</i>	40
<i>Figura 9. Tres servidores por separado</i>	41
<i>Figura 10. Capas del modelo OSI</i>	49
<i>Figura 11. Cabeceras OSI anidadas</i>	49
<i>Figura 12. Capas y cabeceras del modelo de Internet</i>	51
<i>Figura 13. Cabeceras IPv4 y IPv6</i>	61
<i>Figura 14. Series de tiempo que representan el tráfico de voz y datos sobre una red local</i>	79
<i>Figura 15. Espacio CASUAL y sus planos de agregación</i>	82
<i>Figura 16. Interpretación del espacio CASUAL</i>	83
<i>Figura 17. Niveles On-Off del modelo CASUAL y sus distribuciones probabilísticas</i>	87
<i>Figura 18. Distintos tipos de distribución binomial</i>	88
<i>Figura 19. Funciones de densidad gamma para distintos pares de parámetros α y β</i>	89
<i>Figura 20. Variación del índice de cola pesada</i>	91
<i>Figura 21. Niveles del modelo CASUAL y los pares de estados On-Off</i>	94
<i>Figura 22. Niveles On-Off del modelo CASUAL en el eje de servicios</i>	95
<i>Figura 23. Interpretación de los parámetros en el nivel de sesión</i>	104
<i>Figura 24. Interpretación del nivel de ráfaga</i>	105



ÍNDICE DE TABLAS

<i>Tabla 1. Clasificación de algunos modelos aplicados a la medición de Teletráfico.....</i>	<i>34</i>
<i>Tabla 2. RFC 2045, Algunos tipos y subtipos MIME</i>	<i>44</i>
<i>Tabla 3. Comparativa entre el modelo OSI y el de Internet.....</i>	<i>55</i>
<i>Tabla 4. Protocolos comunes en Internet</i>	<i>58</i>
<i>Tabla 5. Algunos formatos de imágenes y sus características.....</i>	<i>66</i>
<i>Tabla 6. Tipos de objetos principales: características y parámetros.....</i>	<i>69</i>
<i>Tabla 7. Distintas tecnologías de acceso.....</i>	<i>72</i>
<i>Tabla 8. Participación de suscriptores por segmentos en Colombia</i>	<i>75</i>
<i>Tabla 9. Distribución de suscriptores de Internet por tipo de acceso, 2008</i>	<i>76</i>
<i>Tabla 10. Cantidad de suscriptores según rango de velocidad</i>	<i>76</i>
<i>Tabla 11. Capacidad demandada al descargar una página web.....</i>	<i>106</i>
<i>Tabla 12. Capacidad requerida por páginas web en Colombia.....</i>	<i>106</i>
<i>Tabla 13. Capacidad con usuarios del lado del servidor.....</i>	<i>108</i>
<i>Tabla 14. Capacidad requerida para el servicio web usando los parámetros de Facebook</i>	<i>109</i>
<i>Tabla 15. Capacidad requerida para atender el servicio web en cada uno de los países.....</i>	<i>110</i>
<i>Tabla 16. Capacidad total demandada</i>	<i>111</i>
<i>Tabla 17. Capacidad requerida según el tipo de acceso en el modelo CASUAL</i>	<i>112</i>
<i>Tabla 18. Capacidad requerida en FBM y CASUAL.....</i>	<i>114</i>

ÍNDICE DE ACRÓNIMOS

A

ADSL (Asymmetric Digital Subscriber Line), 31
AR (AutoRegresivo), 29
ARQ (Automatic Repeat-reQuest), 62
ASP (Active Server Pages), 38
ATM (Asynchronous Transfer Mode), 31

B

BMP (BitMaP), 68
BRI (*Basic Rate Interfaces*), 72

C

CAIDA (Cooperative Association for Internet Data Analysis), 22
CASUAL (*Cubo de Acceso, Servicio y Usuario de Asignación Libre*), 8
CASUAL (Cubo de Acceso, Servicios y Usuarios de Asignación Libre), 16
CDNs (Content Delivery Network), 107
CGI (Common Gateway Interface), 66
COM (*Connection Oriented Model*), 27
CSS (Cascading Style Sheets), 42
CTC (Copenhagen Telephone Company), 22

D

DHCP (Dynamic Host Control Protocol), 63
DHTML (Dynamic HTML), 66
DNS (*Domain Name System*), 42
DSL (Digital Subscriber Line), 72
DVD (Digital Versatile Disc), 68

E

EDGE (Enhanced Data Rates for GSM Evolution), 74

F

FTP (File Transfer Protocol), 17
FTTB (Fiber-To-The-Building), 73
FTTC (Fiber-To-The-Curb), 73
FTTD (Fiber-To-The-Desk), 73
FTTN (Fiber-To-The-Node), 73
FTTP (Fiber-To-The Premises), 73

G

GIF (Graphics Interchange Format), 43
GPRS (Packet Radio Service), 74
GSM (Global System Mobile), 73

H

HDTV (High Definition TV), 68
HSCSD (High Speed Circuit Switched Data), 74
HSDPA (High Speed Downlink Packet Access), 74
HSPA (High-Speed Packet Access), 74
HTML (HyperText Markup Language), 41
HTTP (HyperText Transfer Protocol), 31

I

IAX (Inter Asterisk eXchange), 69
ICMP (Internet Control Message Protocol), 60
IEEE (Institute of Electrical and Electronics Engineers), 73
IGMP (Internet Group Management Protocol), 60
IID (*Independiente e Idénticamente Distribuida*), 28
IMAP (Internet Message Access Protocol), 56
IP (*Internet Protocol*), 16
ISDN (Integrated Services Digital Network), 72
ISMA (Internet Statistics and Metrics Analysis), 22
ISO (International Organization for Standardization), 49
ISP (Internet Service Provider), 19
ITU (*International Telecommunication Union*), 82
ITU (International Telecommunications Union), 27

J

JPEG (Joint Photographic Experts Group), 43
JSP (JavaServer Pages), 39

L

LAN (Local Area Network), 56
LMDS (Local Multipoint Distribution Service), 73
LRD (*Long Range Dependence*), 25
LTE (Long Term Evolution), 74

M

MA (Moving Average), 29
MGCP (Media Gateway Control Protocol), 69
MIDI (Musical Instrument Digital Interface), 69
MIME (Multipurpose Internet Mail Extensions), 43
MMS (Multimedia Message Service), 32
MP3 (MPeg audio layer 3), 69
MPEG (Moving Picture Experts Group), 43
MPLS (Multi-Protocol Label Switching), 26

N

Nc (Nivel de conexión), 108
NIMI (National Internet Measurement Infrastructure), 22
NNTP (Network News Transport Protocol), 41
Nr (Nivel de ráfaga), 108
Ns (Nivel de sesión), 108
NSFNET (National Science Fundation Network), 22

O

OSI (Open System Interconnection), 49

P

P2P (Peer-To-Peer), 17, 75
PDF (Portable Document Format), 43
PHP (Personal Home Page, Hypertext Pre-processor), 38
PNG (Portable Network Graphics), 67
POM (Page Oriented Model), 26
POP (Post Office Protocol), 41, 56
PRI (Primari Rate Interfaces), 72
PSTN (Public Switched Telephone Network), 72

Q

QoS (Quality of Service), 107

R

RA (Real Audio), 69
RAM (Random Access Memory), 41
RIP (Routing Information Protocol), 63

RLC (Run Length Code), 68
RTT (Round-Trip delay Time), 45

S

SCCP (Skinny Client Control Protocol), 69
SIP (Session Initiation Protocol), 69
SMTP (Simple Mail Transfer Protocol), 41
SNMP (Simple Network Management Protocol), 63
SRD (Short Range Dependence), 23, 26

T

TCP (Transmission Control Protocol), 31
TCP/IP (Transmission Control Protocol / Internet Protocol), 27
TES (Transform- Expand-Sample), 23
TES (Transform-Expand-Sample), 25
TLC (Teorema del Límite Central), 94
TOM (Time Oriented Model), 27

U

UDP (User Datagram Protocol), 60
UMTS (Universal Mobile Telecommunications System), 74
URLs (Localizadores Uniformes de Recursos), 42
UTP (Unshielded Twisted Pair), 56

V

VCD (Video Compact Disc), 68
VoD (Video on Demand), 31, 73
VoIP (Voice over Internet Protocol), 82

W

W3C (WWW Consortium), 22
WAV (Waverform Audio File), 69
WCA (Web Characterization Activity), 22
WCG (Web Characterization Group), 22
WiFi (Wireless Fidelity), 73
WiMax (Worldwide Interoperability for Microwave Access), 73
WSAES (World Scientific and Engineering Academy and Society), 81

b - bit

B - Byte (1 B = 8 b),

kB¹ - kilobyte (1 kB = 10³ B = 8 * 10³ b)

MB - megabyte (1MB = 10⁶B = 8 * 10⁶b)

GB - Gigabyte (1GB = 10⁹B = 8 * 10⁹b)

s - segundo

ms - milisegundo (1ms = 10³s)

μs - microsegundo (1μs = 10⁶s)

bps - bit por segundo

kbps - kilobit por segundo (1 kb/s = 10³ b/s)

Mbps - megabit por segundo (1Mb/s = 10⁶b/s)

Gbps - Gigabit por segundo (1 Gb/s = 10⁹ b/s)

Hz - hertz

kHz - kilohertz (1kHz = 10³ Hz)

MHz - megahertz (1MHz = 10⁶Hz)



¹ kB se ha utilizado históricamente para hacer referencia tanto a 1024 (2¹⁰) bytes como a 1000 (10³) bytes, todo depende el contexto

INTRODUCCIÓN

Una pregunta es suficiente para comenzar un camino, para construir, para dar un paso más en un sendero desconocido, y eso es precisamente lo que ocurrió con este proyecto. En el periodo de Mayo-2010 a Enero-2011, tiempo en el que trabajé como practicante en el *Departamento de Planeación y Desarrollo de Negocios* de la Empresa *Internexa S.A. E.S.P.* surge la pregunta ¿Es posible a partir de la demanda de los contenidos de Internet en un país, estimar la demanda de tráfico IP (*Internet Protocol*)? Mi respuesta fue sí, pero en ese momento sólo existían en mi mente montón de ideas difusas de cómo funcionaba Internet y sus contenidos.

Una rápida investigación sobre los temas relevantes del Teletráfico, permitió identificar el modelo **CASUAL** (*Cubo de Acceso, Servicios y Usuarios de Asignación Libre*) como un firme candidato para la solución de la pregunta. Se hizo una primera implementación del modelo, adaptando algunos de sus parámetros a las exigencias impuestas por la base de datos adquirida. Se obtuvieron resultados satisfactorios y la inquietud de conocer más al detalle el funcionamiento de la Red, las aplicaciones, los contenidos y cómo influyen en el tráfico demandado por los usuarios.

Estas inquietudes motivaron la profundización en el tema a través de tres áreas del conocimiento y que soportan este proyecto: las Estadísticas Descriptivas, la Teoría del Teletráfico y la Comunicación de Datos. Estas se amalgaman de forma estructural en el desarrollo de este texto y justifican la adaptación que se hace a algunos parámetros del modelo **CASUAL** para su uso en un ambiente administrativo, reduciendo los costos de equipos de medición especializados.

Dicha adaptación resulta interesante, porque la relación entre los generadores de contenidos, operadores de internet, y usuarios, obliga en el caso de los operadores de Internet, a la planificación y diseño de redes de comunicación que se conecten e interconecten entre sí, de la forma más eficaz posible, para mantener el movimiento

continuo de la curva de oferta y demanda de contenidos, esperando que esto redunde en la demanda de ancho de banda.

Lo anterior es uno de los propósitos de esta investigación, entender como el crecimiento en la demanda de contenidos de Internet, impacta en la capacidad de la red disponible. Con este fin, se revisan distintos modelos utilizados comúnmente para el cálculo de tráfico de una red de Internet. Se resalta el modelo **CASUAL** y se implementa para el caso de descarga y visualización de páginas web, luego se evalúa el desempeño respecto a un modelo tradicional.

Para poder llevar a cabo lo anterior, se ha realizado los siguientes supuestos: se espera que si hay contenido, este tiene un tamaño cuya unidad es el byte, y que viaja a través de la red desde los servidores hasta los usuarios, por tanto debe ser posible conocer el tamaño de la red necesario para soportar este flujo de información de una manera eficiente, sin importar su topología. Si se considera un flujo permanente, entonces, las variaciones de este flujo dependerán básicamente del tipo de servicios al que accedan los usuarios, que pueden ser Web, **P2P (Peer-To-Peer)**, Stream, Video y Música online, **FTP (File Transfer Protocol)**, emails, voz y juegos. Aunque el flujo es permanente, su naturaleza es discreta debido al empaquetamiento de la información, y por tanto hay momentos de presencia y ausencia de información, una analogía que puede ayudar a familiarizar el concepto puede ser los vagones de un tren cuyo contenido es información pero entre cada vagón hay una distancia o ausencia de información, queda implícito que el tamaño de la carga que transporta cada vagón también es diferente.

La investigación se centra dentro del marco del servicio Web, y el enfoque básico usado para la adquisición de los datos, está basado en estadísticas globales como suscriptores de internet, páginas más visitadas, tiempo total de descarga, número total de objetos descargados, tamaño total de la descarga y otros, con el objetivo de orientar el modelo al enfoque administrativo.

En la metodología de desarrollo se han utilizado las recomendaciones de la Unión Internacional de Telecomunicaciones, que sugiere para el cálculo de la demanda, tres pasos, la medición, la modelización y la predicción. Para la medición, los datos son tomados de publicaciones online de sitios web reconocidos por ese tipo de servicios y



cuentan con una alta fiabilidad. Para la modelización, se hace revisión de un grupo de modelos utilizados convencionalmente para el cálculo del tráfico en Internet, pero se utiliza el modelo **CASUAL** como más adecuado según la naturaleza de los datos y la predicción se limita sólo al eje de usuarios, usando un modelo autoregresivo, se aclara que cada uno de los ejes (eje de accesos y eje de servicios) puede usar un modelo de predicción independiente. En este caso sólo se usará un modelo en el eje de usuario. Todo esto para adaptar un modelo de tráfico que ayude a dimensionar y predecir el ancho de banda disponible, con el fin de desarrollar herramientas, dirigidas a la planificación estratégica de la capacidad para un oportuno desarrollo de redes de comunicación de banda ancha.

Internet ha transformado altamente el mundo en poco tiempo, la web ha roto fronteras y abierto un abanico de opciones. Las aplicaciones informáticas se han expandido a todos los niveles de la sociedad acercándose a un mayor número de personas, a su quehacer diario, facilitando la generación de contenido, redes sociales, comercio electrónico, video llamadas entre otros. Los gobiernos y operadores de redes realizan un esfuerzo para que las personas tengan acceso a Internet, en virtud del progreso cultural y tecnológico. Ellos deben garantizar que este crecimiento sea consecuente con la demanda de contenidos y aplicaciones que hacen los usuarios. En este sentido el trabajo realizado aquí se convierte en una herramienta para estimar dichos requisitos.

Convenciones tipográficas

Con el objetivo de hacer más comprensible el texto y evitar ambigüedades en la interpretación de las ideas contenidas aquí, se hace uso de distintas convenciones tipográficas que se utilizarán a lo largo del documento:

En las referencias bibliográficas citadas dentro del texto, se usará la fuente Eras Light ITC, con estilo cursivo entre comillas dobles acompañadas al final con un número entre paréntesis que identifica al documento citado al final en la bibliografía, ej: *“que el modelo se parezca al tráfico real, que sea fácil de estimar sus parámetros y que los parámetros tengan*



significado físico" (1). Cuando sólo se esté haciendo mención a la referencia como fuente de consulta se colocará únicamente el número entre paréntesis.

Los acrónimos y siglas definidos en el índice de acrónimos se muestran en negrita y siempre que se mencionen por primera vez en el cuerpo del texto vendrán acompañados por el significado de la sigla entre paréntesis y en letra cursiva, resaltando en negrita las letras que pertenecen a la sigla, ej: **ISP** (*Internet Service Provider*)

Las palabras escritas en un idioma distinto al español se escribirán en cursiva, ej: *browser*.

Cuando se hace referencia a un programa, o página web se colocarán en negrita las interfaces que aparecen en pantalla y con el signo mayor que ">" se indicará la secuencia de acciones que debe seguirse para interactuar con la página o programa, ej: **www.alexa.com > Top sites > By country > Colombia**

Las fórmulas y ecuaciones se presentan en cursiva y con tipo de letra *Cambria Math*, ej:
$$X(t) = mt + \sqrt{a}Z(t)$$

Los nombres de empresas y compañías se muestran con Mayúsculas inicial y letra cursiva, ej: *Internexa, Microsoft, Telegeography, etc.*

Los nombres de los programas y páginas web se muestran con el tipo de letra (Kosuka Mincho pro L). Ejemplo: Microsoft, Paint, Firefox, Matlab, etc.

Las extensiones de los ficheros, y el código incluido dentro del texto se muestra con un tipo de letra Courier. Ejemplo: .html, .jpg, .gif, programamejorado = "Elche", etc.

OBJETIVOS

Objetivo general

Estimar el tráfico total generado en algunos países de la región suramericana por la descarga y visualización de los contenidos de páginas web, mediante la adaptación e implementación de un modelo de tráfico, el modelo **CASUAL**.

Objetivos específicos

- Revisar los distintos modelos utilizados para el cálculo de tráfico en una red de internet
- Adaptar el modelo **CASUAL** en el caso de descarga y visualización de páginas web, usando datos globales de uso de Internet para este tipo de servicio
- Evaluar el desempeño del modelo escogido respecto a un modelo tradicional

UNIVERSIDAD
DE ANTIOQUIA
1803



1. MODELOS DE TELETRÁFICO

En esta sección se hace un recuento de distintos modelos utilizados en la teoría de Teletráfico para la medición de Internet, siguiendo tres criterios básicos: *“que el modelo se parezca al tráfico real, que sea fácil de estimar sus parámetros y que los parámetros tengan significado físico”* (1), luego se presentan las características principales que los identifican y por último se cuenta un poco de que trata cada modelo.

1.1. Breve Historia

Antes de comenzar, es importante entender la teoría del teletráfico *“como la aplicación de la teoría de las probabilidades a la solución de problemas concernientes a la planificación, evaluación de la calidad, operación y mantenimiento de los sistemas de comunicación”* (2). Esta teoría es amplia en sus cimientos matemáticos y puede ser aplicable a otros sistemas como tráfico aéreo, bandas transportadoras, gestión de almacenamiento y todo sistema que implique servicios. En este sentido la teoría de teletráfico se puede entender como una disciplina para la planificación de medios, procesos y operaciones, con un objetivo específico: *“Lograr calcular el tráfico en unidades bien definidas mediante modelos matemáticos y determinar la relación existente entre la calidad de servicio y capacidad del sistema de tal manera que la teoría se convierta en una herramienta útil para la planificación de inversiones”* (2).

Tiene su origen después de la invención de las líneas telefónicas, cuando surge la necesidad de conocer el número de líneas necesarias para conectar dos ciudades, lo que planteó el reto de conocer el comportamiento de los usuarios, modelar el servicio de telefonía tradicional y dimensionar los recursos mínimos para mantener cierto grado de calidad en el servicio ofrecido. En 1909, con el artículo *“The Theory of Probabilities and Telephone Conversations”* Agner Krarup Erlang resuelve el problema demostrando que la distribución de Poisson era aplicable al tráfico telefónico, tiempo después en 1917, resuelve el problema de los tiempos de espera. En 1948, 20 años después de su



muerte, la **CTC** (*Copenhagen Telephone Company*) publica la recopilación de todo su trabajo (3).

Desde su origen esta teoría estuvo orientada a resolver problemas de telecomunicaciones, a medida que los sistemas digitales aparecieron también se ha usado para resolver los problemas de transmisión de datos. La aparición de Internet y su crecimiento acelerado hizo que a principios de la década de los 90s que comunidades de investigadores e **ISPs** sumaran esfuerzo para determinar cuáles parámetros deberían tenerse en cuenta, y desarrollar modelos que se ajustaran a la realidad de Internet, “Estos incluyen las métricas de rendimiento tales como latencia, paquetes, pérdida de paquetes, utilización del enlace, calidad de servicio y tiempos de protocolo de enrutamiento de convergencia” (4).

Sobresalieron entre estas comunidades de investigadores la **NSFNET** (*National Science Foundation Network*) que fue disuelta en 1995 y que estaba encargada de recolectar datos históricos del tráfico de Internet (5). La **NIMI** (*National Internet Measurement Infrastructure*) un proyecto de infraestructura de medidas distribuido (6). Para 1998 comienza un proyecto para desarrollar la caracterización de Internet **WCA** (*Web Characterization Activity*) (5), que se convierte en la **WCG** (*Web Characterization Group*) de la **W3C** (*WWW Consortium*). En 1993 nace **CAIDA** (*Cooperative Association for Internet Data Analysis*) que, a partir de 1996 hasta hoy, se ha dado a conocer organizando pequeños simposios con el fin de trabajar en la recolección y análisis de estadísticas globales de tráfico y prestaciones de Internet, conocido como **ISMA** (*Internet Statistics and Metrics Analysis*) (7).

A partir del año 2000 han proliferado los grupos de investigación interesados en la teoría de teletráfico enfocados al desarrollo y entendimiento de Internet, dando a lugar al mejoramiento y aparición de distintos modelos que permiten comprender el tráfico generado por los distintos servicios que operan sobre la Red.

1.2. Tipos de modelos de Teletráfico

Dependiendo del tipo de tráfico que circula por Internet y los distintos servicios que este soporta, han aparecido diferentes modelos utilizados para representar, analizar y



comprender dichos tráficos. Esto ha dado origen a múltiples formas de abordar los problemas relacionadas con el tráfico de la Red, surgiendo así una manera de clasificar los modelos, como macroscópicos, microscópicos, conductistas, estructuralistas, deterministas, estadísticos, analíticos, los basados en cadenas de Markov, los basados en tráfico de fluidos, los autorregresivos, los autosimilares, los TES (*Transform-Expand-Sample*), los de renovación de tráfico, los no correlacionados, los correlacionados con SRD (*Short Range Dependence*), los correlacionados con LRD (*Long Range Dependence*), los modelos cerrados, los modelos activos, los modelos pasivos, los modelos caóticos, los de grandes desvíos y en particular para el tráfico web, los orientados a páginas, los orientados a tiempos y los orientados a conexión. Con esta clasificación se estaría abarcando la mayor cantidad de los modelos de teletráfico existentes por lo que no es necesario extenderla más, a continuación se explicará cada uno de ellos:

Los *modelos macroscópicos*, llamados así por el tipo de medidas que toman, total de bytes transferidos, crecimiento de la web, servicios más usados, tipos de datos transferidos, número de suscriptores, etc. Este tipo de modelos consideran que el cuello de botella está en los extremos, servidores y clientes, los primeros deben atender una multitud de peticiones y los segundos tienen unas limitaciones de ancho de banda.

Los *modelos microscópicos*, o también conocidos como modelos de capa de aplicación y capa de enlace, toman medida del número de conexiones abiertas, duración de la conexión, tiempo entre conexiones, conexiones ascendentes, conexiones descendentes, duración de la sesión, número de ráfagas, etc. Este tipo de modelos consideran que los cuellos de botella se encuentran en los distintos niveles de agregación de tráfico.

Los *modelos conductistas*, intentan describir los procesos con el mínimo de parámetros posible, y que estos tengan un significado físico, aunque en todos los casos no es posible; por eso, son conocidos como modelos de caja negra, porque buscan imitar las propiedades estadísticas del tráfico sin abordar las estructuras internas que generan el mismo (5).



Los *modelos estructuralistas*, al contrario de los conductistas, intentan describir de la forma más exacta posible la estructura interna de los sistemas que generan el tráfico. Son extremadamente detallados y complicados desde el análisis matemático y algunos analíticamente intratables (5).

Los *modelos deterministas*, intentan describir el peor de los casos, lo que supone una sobre estimación de los recursos que aseguran el tráfico. Los parámetros de estos modelos son fáciles de hallar; sin embargo, es necesario conocer con anterioridad los patrones de tráfico (8).

Los *modelos estadísticos*, utilizan aproximaciones para simular las fuentes de tráfico, pero es necesario que cada fuente sea representada mediante un modelo analítico. Este tipo de modelos exige ser ajustado constantemente, corrigiendo los valores de sus parámetros. Son extremadamente complicados, pues la actualización de los parámetros es difícil, mucho más cuando se pretende hacer en tiempo real (8).

Los *modelos analíticos*, describen la distribución de la variable aleatoria a modelar, son simples, tratables matemáticamente, lacónicos comparados con la realidad, pero si los parámetros que componen el modelo pueden ser obtenidos de forma empírica estos modelos predicen de forma exacta la función de distribución modelada (8). Sin embargo existe un inconveniente *"a menudo se deben introducir aproximaciones para obtener un modelo analítico tratable y esto hace que los valores obtenidos no se acerquen a los valores del sistema original"* (9).

Los *modelos basados en cadenas de Markov*, describen procesos estocásticos, con estados finitos, excluyentes, cuya transición de un estado a otro se producen en periodos uniformes de tiempo y su probabilidad depende del estado anterior.

Los *modelos basados en tráfico de fluidos*, en estos modelos *"el tráfico es tratado como un fluido continuo el cual fluye por los distintos enlaces de la red. Así pues, la agregación es consecuencia de la agregación de fluido generado por un conjunto de fuentes On-Off independientes"* (10), en otras palabras *"estos modelos consideran procesos en los que el número de observaciones es tan grande que la información que aporta cada una de ellas es insignificante, pudiendo ser tratadas todas ellas como un flujo constante en el que solamente los cambios determinan la variación de los parámetros de dicho flujo. Es*



precisamente la señalización de dichos cambios la que lo modela. Es por ello que estos modelos son utilizados bien ante aplicaciones y servicios de tasa constante, donde dichos parámetros son también prácticamente constantes como es el caso de ATM, o en las transferencias de datos que llevan a cabo algunas aplicaciones TCP/IP" (8).

Los *modelos autorregresivos*, conocidos también como estocásticos lineales, definen una variable aleatoria como una función de las variables aleatorias anteriores en un intervalo de tiempo que incluye determinado número de esas variables, originando una distribución marginal con dependencia de corto plazo (11).

Los *Modelos autosimilares*, describen los procesos que al cambiar su escala de observación su comportamiento estadístico de la autocorrelación sigue siendo el mismo, de ellos se dice que son fractales en el tiempo, pues su autocorrelación se mantiene a distintas escalas de tiempo (12). También son conocidos como modelos correlacionados con **LRD (Long Range Dependence)**.

Los TES (Transform-Expand-Sample), "los cuales obtienen una función de distribución aproximada mediante la transformación de una secuencia uniformemente distribuida (mediante el uso de histogramas). Su capacidad para capturar las principales características de la función de autocorrelación de la secuencia original, les ha hecho interesantes para su uso en modelos específicos de tráfico de VBR (Variable Bit Rate), Ethernet y Web" (8). Es una metodología que supone que la serie de tiempo de tráfico está disponible en todo momento, lo que permite tomar los datos y hacer los cálculos de autocorrelación al mismo tiempo (11).

Los *modelos de renovación de tráfico*, también conocidos como modelos no correlacionados o modelos de tráfico sin memoria, consideran las variables del sistemas independientes, sin conocimiento de eventos anteriores y uniformemente distribuidas (11).

Los *modelos no correlacionados*, supone que el tráfico agregado de conexiones, paquetes, llamadas, flujos que ingresa a la red debido al uso que hacen usuarios independientes, también es independiente, de tal manera que los tiempos de demanda en los nodos de red no se correlacionan y por tanto entre ellos no existe dependencia de largo plazo o de rango largo (13).



Los *correlacionados con SRD (Short Range Dependence)*, son para redes de múltiples servicios, donde los tiempos de llegadas de los paquetes incluyen correlaciones que decaen en forma exponencial y rápidamente en el tiempo permitiendo de esta forma la tratabilidad matemática (13).

Los *modelos cerrados*, modelos que utilizan la teoría de control óptimo para resolver problemas de enrutamiento, el enrutamiento de redes de circuitos virtuales, conmutación por etiquetas **MPLS (Multi-Protocol Label Switching)** y otros. Conocidos con ese nombre porque su estructura de control es un lazo cerrado (13).

Los *modelos activos*, son aquellos que para inferir el retardo, las pérdidas, capacidad efectiva entre otros, introducen tráfico para las mediciones en la red que pretenden modelar, este tipo de modelo incrementan la carga de la red e interfieren en su desempeño pero a cambio obtienen valores más cercanos a las características que pretenden simular (13).

Los *modelos pasivos*, son aquellos que parten del tráfico capturado sobre la red, centrándose en las características visibles de dicho tráfico, lo que no permite inferencias exactas, sino aproximadas. Esta limitación obliga a este tipo de modelos a la sencillez y simplicidad (13).

Los *modelos caóticos*, son aquellos que permiten analizar trazas de tráfico con alta correlación y fractalidad, en general estas trazas provienen de sistemas dinámicos, no lineales. El grado de fractalidad de este tipo de sistema está asociado al exponente de Lyapunov, y sirven para analizar características de desempeño, en especial los transientes de los sistemas de colas (13).

Para el tráfico servicio web, la clasificación se puede hacer de la siguiente forma:

Los *modelos orientados a páginas POM (Page Oriented Model)*, estos modelos se centran en el nivel de sesión, página y paquete. Son modelos activos que inyectan paquetes en los distintos enlaces. No analizan el nivel de conexión porque es en el de página donde se concentra el mayor número de ráfagas (5).



Los *orientados a tiempos TOM (Time Oriented Model)*, considerados así porque el parámetro referente a la duración de las páginas es dado directamente y no es consecuencia del tamaño de esta. Estos modelos se centran en el nivel de enlace, pero además incluyen el nivel de sesión, página y paquete (5).

Los *orientados a conexión COM (Connection Oriented Model)*, se ubican en el nivel de aplicación. Estos modelos incluyen los niveles de sesión, página y conexión, son prácticos porque separa la influencia que tienen los usuarios, la red y los protocolos en el tráfico. Este tipo de modelos se utilizan para analizar el tráfico **TCP/IP (Transmission Control Protocol / Internet Protocol)** que cruza la Red (5).

los *modelos de grandes desvíos*, usados para modelar dos situaciones en los nodos de las redes: la primera, el comportamiento de los enlaces si el buffer es grande con lo que se obtiene la distribución del retardo y la probabilidad de pérdida de los paquetes, además de la capacidad para atender simultaneidad de distintas ráfagas. La segunda, cuando el enlace es alimentado por muchas fuentes, para saber la capacidad de desempeño ante la agregación de cada una de ellas, la calidad de servicio y la disposición de los recursos (14).

Con esta breve descripción, se intentó mostrar los grupos más importantes, donde se encuentran la mayoría de los modelos existentes, sin embargo, es bueno aclarar que no se ha expuesto el conjunto total de clasificación de modelos, hicieron falta por ejemplo, los modelos sin memoria, los modelos de equilibrio, los modelos abiertos, los modelos basados en teoría de juegos, los no correlacionados con **LRD** y otros, que no han sido explicados por considerarlos de una u otra forma, subconjuntos de los anteriores.

1.3. Algunos modelos usados en teletráfico

Una parte importante de la teoría de teletráfico es la caracterización de la demanda de tráfico, que según la recomendación E-490.1 (15) de la **ITU (International Telecommunications Union)** se compone de tres partes: la modelización del tráfico, la medición de tráfico y la predicción de tráfico. Según esto, *“la caracterización de tráfico se efectúa por medio de modelos que se aproximan al comportamiento estadístico de tráfico de red en una gran población de usuarios. Los modelos de tráfico adoptan hipótesis*



simplificadas referentes a los procesos de tráficos complicados” (2). Se entiende entonces que un modelo de tráfico es “una abstracción matemática más o menos compleja que trata de imitar alguna o varias características estadísticas de un tipo de tráfico real o de un flujo concreto en particular” (16).

A continuación se enunciarán algunos modelos usados en teletráfico tanto en la parte de modelización como en la parte de predicción:

1.3.1 Modelos de Poisson

Se caracterizan porque hacen uso de los procesos de Poisson y de ahí su nombre, en ellos la función de distribución de probabilidad $X(t_i)$ es IID (*Independiente e Idénticamente Distribuida*) para el valor de la variable aleatoria en cada t_i , esto sin importar su distribución por lo que su autocorrelación es cero, esta característica se conoce como renovación. Modelan la llegada de paquetes al servidor o al cliente, separando el intervalo de tiempo entre llegadas de forma proporcional con incrementos estacionarios. La ventaja de este tipo de modelos es que la media λ , es su parámetro principal, cuya información sirve para estimar los valores de alta utilización o ancho de banda promedio en las redes. Otra característica importante es que este tipo de procesos de Poisson la sucesión de las variables aleatorias tienen una distribución exponencial (17).

1.3.2 Modelos de Bernoulli

Reciben su nombre porque hacen uso de los procesos de Bernoulli, que son una particularización de los procesos de Poisson en el dominio discreto de tiempo y al igual que estos, también presentan la característica de renovación, donde la función de distribución de probabilidad $X(t_i)$ es IID para el valor de la variable aleatoria en cada t_i , y sin importar su distribución, su autocorrelación es cero. Modelan la llegada de paquetes al servidor o al cliente, mediante una función de distribución binomial, donde cada llegada es independiente de la anterior. Sirven para modelar ráfagas de tipo *On-Off* (*On-Llegada, Off-Espera*) cuyo estado *Off* se modela con una distribución geométrica. El parámetro principal en este tipo de modelos es la probabilidad de llegada p (8).



1.3.3 Modelo tipo fase

Una de las características principales de este modelo es que presentan varios o un estado absorbente de Markov, de tal manera que si un proceso entra en dicho estado permanecerá en el indefinidamente y la probabilidad de pasar a otro estado se hace nula, interpretándose este fenómeno como pérdida de memoria. Lo que se busca es que el proceso no llegue al estado absorbente, para estos se condiciona la variable aleatoria a determinadas distribuciones como la exponencial, la Erlang u otra, de tal manera que la probabilidad de pasar de un estado a otro queda determinada por el tipo de distribución. Los parámetros principales de este modelo son los estados no absorbentes d , y una matriz de intensidades, similar a una matriz de transición pero cuyos valores dependen del tipo de distribución (17). Este modelo se usa para la congestión en los nodos y pérdidas de paquetes.

1.3.4 Modelo lineal autoregresivo

En este modelo el valor actual de la serie depende linealmente de los valores anteriores. Este modelo sigue un proceso **AR(p)** (*AutoRegresivo de orden p*) de la forma $X_n = a_0 + \sum_{r=1}^p a_r X_{n-r} + \varepsilon_r$, $n > 0$. Donde X_n representa el tiempo de llegada entre los paquetes n y los paquetes $n - 1$. De esta forma el modelo considera la autocorrelación del proceso de llegadas. Cuando se hace la suma ponderada de los valores anteriores que ha tomado la variable aleatoria se deriva a un proceso de **MA** (*Moving Average*) (18).

1.3.5 Modelo de Equilibrio de Wardrop

Este es un modelo de asignación de equilibrio basado en el usuario, de tal manera que *“los tiempos de viaje en todas las rutas utilizadas son menores o iguales que los que requeriría un usuario en cualquier ruta no utilizada”* (19). En otras palabras, si ningún usuario (paquete) del sistema no puede reducir unilateralmente sus tiempos o costos de viaje a través de la red cambiando su ruta, entonces el sistema se encuentra equilibrado. De tal manera que una red equilibrada es aquella donde existen un conjunto de vectores entre el origen y el destino que satisfacen las restricciones de continuidad y la demanda de los usuarios por disminuir su propio tiempo.



1.3.5 Procesos TES

Parten de una serie empírica estacionaria de datos, resultado de una medición de tráfico. A partir de estos datos se siguen tres pasos para cumplir con los requerimientos de fidelidad: Primero, que el histograma de frecuencia de los datos generados sea similar al de los datos originales, de tal manera que se conserve la distribución marginal de la muestra. Segundo, que las medidas estadísticas de los datos obtenidos sean similares al de los datos originales, de tal manera que la autocorrelación tenga el menor retraso posible respecto a la autocorrelación de la muestra y por último, los datos deben pasar la prueba por décima de hipótesis y ser una representación de los datos originales (20).

1.3.6 Modelo movimiento fraccional browniano

Este es el modelo más importante que representa los procesos autosimilares y al ruido gaussiano autosimilar. De tal manera que $Y(t)$ con $t \in R$, es un movimiento fraccional browniano si su parámetro de Hurst H cumple con la condición $0 < H < 1$, pero además tendría características LRD, propias de Internet, si dicho parámetro estuviera en el intervalo $0.5 < H < 1$. La variable aleatoria $Y(t)$ es gaussiana exactamente autosimilar con incrementos independientes, entonces se conoce como ruido gaussiano autosimilar. Cuando las dos características mencionadas se combinan, se obtiene un proceso no correlacionado con LDR (21).

1.3.7 Modelos On-Off

Este modelo se puede usar para modelar fuentes de tráfico que llegan a una red, siempre y cuando haya un periodo activo de transferencia de datos llamado *On* y un periodo de espera o *Thinking Time* llamado *Off* (Ver **figura 1**). De esta forma, el tráfico considerado alterna entre periodos *On* y periodos *Off* cuyas longitudes son independientes y pueden tener distribuciones distintas en un intervalo de tiempo cualquiera. Sin embargo, cuando se genera tráfico, periodo activo *On*, este se genera a una velocidad r y puede ser representado mediante una distribución de probabilidad, generalmente una binomial para el caso de tráfico **HTTP** (*HyperText Transfer Protocol*). La representación más común para estos modelos es una cadena de Markov de dos estados.



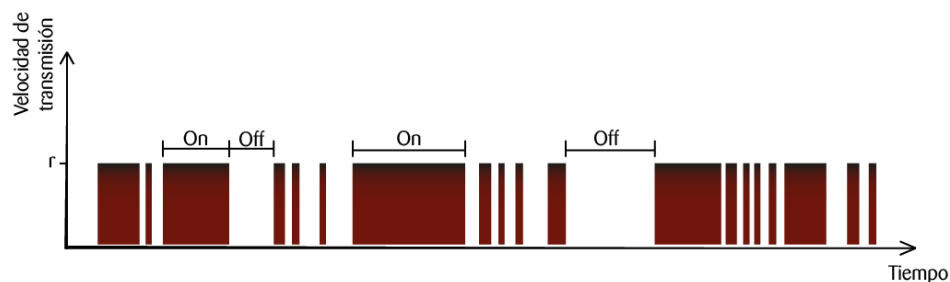


Figura 1. Representación del tráfico On-Off, (nivel de paquetes)

El modelado *On-Off* desde sus inicios ha sido usado para describir distintos tipos de tráfico y tecnologías de red:

En la capa física, que representa la conexión física entre el emisor y el receptor, se ha usado para estimar el tráfico en tecnologías de acceso inalámbrico, Ad-Hoc, redes ópticas y redes telefónicas (22).

En la capa de enlace de datos, encargada de transferir el mensaje de una computadora a la siguiente en la ruta de red entre el emisor y el receptor, se ha utilizado para modelar el tráfico agregado y optimizar los recursos de redes para **ADSL** (*Asymmetric Digital Subscriber Line*), Ethernet y **ATM** (*Asynchronous Transfer Mode*) (22).

En capa de red, encargada de la transferencia de mensajes extremo a extremo, se ha usado para describir el tráfico **IP** con su comportamiento activo y de espera.

En la capa de transporte, encargada del direccionamiento, descomposición de los mensajes, enlazar el software y establecer las conexiones, se ha usado para modelar el protocolo **TCP** (*Transmission Control Protocol*) (22).

En la capa de aplicación, presentación y sesión, se ha usado para reducir el encolamiento en los buffers y mejorar la calidad del servicio.

De lo anterior, es claro que los periodos de *On-Off* van a estar determinados por el tipo de servicio de Internet que esté generando el tráfico, de tal manera que un usuario que utilice el servicio web generará un tráfico con periodos *On-Off* diferentes a un usuario que esté usando el servicio de **VoD** (*Video on Demand*) y un usuario que esté usando **MMS** (*Multimedia Message Service*). En la **figura 2** puede observarse un ejemplo del tráfico generado por el servicio **MMS** (3).

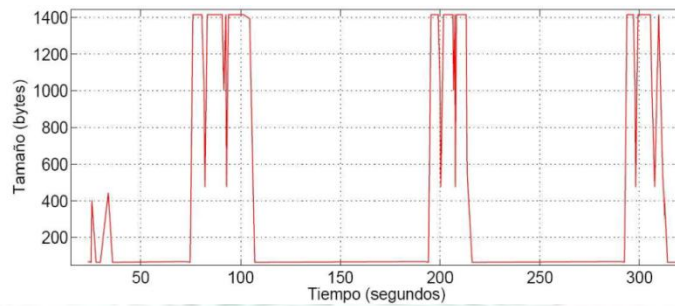


Figura 2. Traza On-Off de tráfico MMS²

Si se analiza el tráfico web a la luz del modelo *On-Off*, un usuario que descarga una página, activa una sesión con una determinada serie de flujos de bajada, luego aparece un periodo de inactividad, donde no hay transferencia de datos debido a que el usuario se encuentra visualizando el contenido que descargó (8).

Por tanto, una sesión puede representarse como una serie de periodos de actividad T_{on} (en segundos) donde se descargan una cantidad de paquetes en bytes. Y una serie de periodos de inactividad T_{off} o $T_{Thinking}$ donde el usuario se encuentra visualizando la página. *“Todos los valores son considerados en valor medio y son diferentes para cada tipo distinto de aplicación existente en la red modelada”* (8). Como se observa en la figura 3.

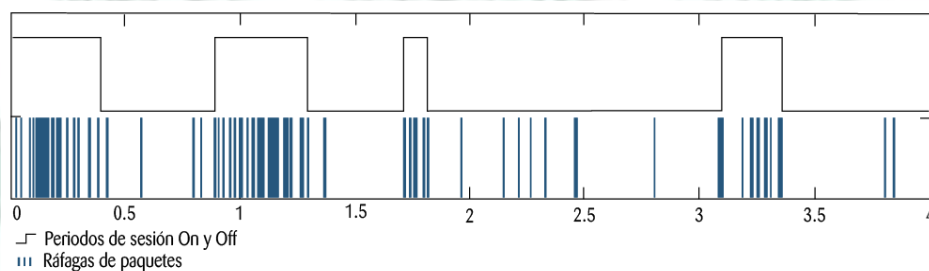


Figura 3. Representación de una sesión en el modelo On-Off³

² Tomado de: academy.itu.int/index.php/events/item/download/667. Modulo I.2 “Técnicas de modelado de tráfico y calidad de experiencia en servicios multimedia” (3)

³ Alzate Monroy, Marco A. "Introducción al tráfico autosimilar en redes de comunicaciones". En Ciencia, Investigación y Desarrollo. Vol. 6. N° 2. pp. 6-17, 2001. (21)

Estos modelos son ideales para analizar redes troncales donde el tráfico agregado corresponde a distintos tipos de servicios, tradicionalmente se había usado para este tipo de análisis distintos niveles de estados *On-Off*, enlazados al estilo de matrioska rusa al interno del estado *On* (8), luego se le han añadido el análisis por multiplexación de múltiples fuentes (22).

En la *Tabla 1* se muestra un resumen de algunos modelos aplicados al Teletráfico, su fecha de aparición y su clasificación según el tipo de modelo descrito en el apartado 1.2.

Es importante aclarar que sólo se resaltan algunas de las características más importantes de cada modelo. Esta es una clasificación subjetiva orientada a mostrar la conveniencia del modelo **CASUAL** para el dimensionamiento de la demanda de tráfico por parte de los usuarios.



Tabla 1. Clasificación de algunos modelos aplicados a la medición de Teletráfico

Fecha	Modelos	Clasificación																						
		COM	TOM	POM	Caóticos	Pasivos	Activos	Cerrados	correlacionados	No correlacionados	Renovación	TES	Autosimilares	Autoregresivos	Fluidos	Markov	Analíticos	Estadísticos	Deterministas	Estructuralistas	Conductistas	Microscópicos	Macroscópicos	
1713	Bernulli						*																	*
1713	Procesos de Poisson																							
s.f.	Tipo Fase																							
1896	Linear Autoregresivo												*											
1903	Media Movil																							
191-	Media Movil Integrada																							
1952	Equilibrio - Wardrop																							
1953	M/M/1 ⁴																*	*						
1953	M/M/1/L ⁵																*	*						
1956	Browniano																	*	*					
1958	Modulados por Markov																	*	*					
1959	Matriz Transición Modulada																	*	*					
1960	Poisson modulado																	*	*					
1988	On-Off																	*	*					
1994	Fraccional Browniano																	*	*					
1997	Trace Sampling																	*	*					
200-	Buffer infinito - AMS																	*	*					
200-	Buffer finito - Tucker																	*	*					
2009	CASUAL																	*	*					*

⁴ Notación de Kendall: "A/B/X/Y/Z/V = Modelo de llegada / Modelo de servicio / Número de servidores / Capacidad del sistema / Disciplina del servicio / Estado del servicio" **M**: Tiempo entre llegadas o servicios exponenciales Markovianos

⁵ Ibid.

2. EJE DE SERVICIOS

En este capítulo se hace una breve exposición de los conceptos que soportan los argumentos y los resultados de este trabajo. Se hace referencia a la comunicación de datos como parte importante del proceso de entendimiento del comportamiento de los niveles de conexión, sesión y ráfaga que usa el modelo **CASUAL** para la estimación de la demanda de tráfico.

Debido al constante y acelerado crecimiento de Internet, las **ISPs** deben encontrar métodos eficientes para calcular la demanda y optimizar los servicios, reduciendo los costos de tiempo y dinero. Para esto existen muchas alternativas, la mayoría de ellas orientadas a equipos de hardware y soluciones de software que hacen mediciones en tiempo real del tráfico que circula por los equipos de red. La conveniencia de este tipo de mediciones es que se realiza directamente sobre los equipos, obteniendo directamente los datos que se necesitan para realizar las estimaciones.

Lo anterior está bien cuando un **ISP** quiere conocer el tráfico que viaja por la red que tiene instalada y es, sin lugar a dudas, la forma óptima de precisar la capacidad de dicha red. Sin embargo, en un país existe más de una empresa **ISP**, cada una con su propia red instalada. Por tanto, si fuera conocida la capacidad de red instalada por cada una de ellas, entonces, la agregación de todas ellas daría como resultado, la capacidad total de la red instalada en un país y sería tarea fácil para una **ISP** determinar la porción de la demanda que está cubriendo. Entonces, la capacidad total de la demanda de Internet en un país, es un dato importante que sirve para orientar la visión del negocio. El conflicto se presenta cuando estos datos no están disponibles al público, y una medición técnica deja de ser la solución.

Todos los temas expuestos en este capítulo, tienen como fin brindarle al lector una visión suficiente, de tal manera que pueda comprender el mecanismo usado para

adaptar los parámetros, necesarios en el cálculo de las probabilidades *On-Off* del modelo CASUAL, de una forma clara y sencilla.

2.1 Sobre la comunicación de datos y las redes

2.1.1 Arquitectura Cliente / Servidor

Se ubica en 1960 el comienzo de la carrera de los sistemas de información. En estos años se usan las líneas telefónicas para la comunicación de datos, sin embargo, el proceso consistía en una computadora central conocida como mainframe que recibía los lotes de mensajes procedentes de una terminal de usuario para procesarlos y luego los enviaba de regreso (23).

Para la década de los 70's aparecen los sistemas de tiempo real y se abandona el esquema de transmisión de información por lote para darle paso al procesamiento individual de transacciones. Aparecen los sistemas de base de datos, y la integración de los sistemas de información al mundo de los negocios (23). Esta época se conoce como el comienzo de Internet, pero no fue hasta la década de los 80's y principios de los 90's que se comienza a usar la arquitectura Cliente/Servidor. Convirtiéndose en la arquitectura propia de Internet.

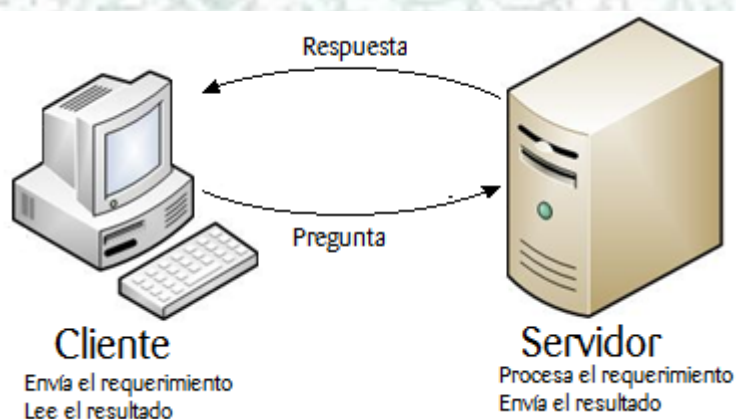


Figura 4. Arquitectura básica Cliente/Servidor

2.1.1.1 ¿Qué es la arquitectura cliente/servidor?

Es una arquitectura flexible y descentralizada, una plataforma que permite enlazar y combinar componentes en los distintos niveles de la red. Separa las aplicaciones entre

las funciones y los datos en el lado del servidor que serán utilizadas por varios usuarios y aquellas en el lado del cliente, asignadas a cada uno de los usuarios (24), ver **figurar 4**.

En el lado del cliente se realizan consultas y solicitudes, interpretación y visualización de la interfaz gráfica, recolección de datos y ejecución de programas, todas ellas acciones de primer plano. Al otro lado, en el servidor, se hace control el acceso a las bases de datos, a los componentes del sistema de información y la administración de redes, esto se conoce como acciones de segundo plano (24).

Aunque las redes están distribuidas alrededor de toda la tierra, el tipo, número de clientes y servidores siguen siendo una cantidad medible. Esta distribución de las redes permiten que los que los usuarios accedan de manera transparente a todos los datos que necesite, en cualquier parte, en cualquier momento, y sin importar como estén almacenados. De hecho para él, también es transparente el sistema de administración de base de datos y la computadora donde estos residen.

Por tanto la arquitectura cliente/servidor es un sistema en el que los procesos o computadores tienen un rol como clientes o como servidores, aunque un mismo computador pueden residir el cliente y el servidor. La ventaja de esta arquitectura es que permite escoger la plataforma más adecuada para cada tarea o proceso. Otras ventajas que se pueden mencionar son:

Los costos: Es económica y ofrece muchas más prestaciones que una computadora central con capacidad para procesar una aplicación, entre estas prestaciones se encuentra migración de contenidos y aplicaciones de un computador a otro con modificaciones mínimas o nulas en cuanto a las instrucciones de los programas (25).

El acceso a la información: El acceso a la información es transparente para el usuario, puede usarla y procesarla a su gusto, puede acceder a ella desde cualquier lugar y momento.

La ergonomía: El usuario no debe preocuparse por la conversión de datos y los diferentes protocolos, el sólo debe interactuar con una interfaz gráfica.



La integración de tecnologías: Conviven sobre la misma red distintas bases de datos, diferentes plataformas propietarias y libres, varios sistemas operativos, etc. Lo que permite y facilita el uso de dispositivos, aplicaciones y software. Además una misma aplicación se puede ejecutar aún si tiene partes distribuidas en la red.

La modularidad y escalabilidad: Permite agregar o eliminar aplicaciones y computadores a la red, dicho de otra forma, permite un crecimiento horizontal o vertical. La escalabilidad o crecimiento horizontal se refiere a la capacidad de suprimir o añadir computadores que hagan uso de una aplicación cliente, sin afectar severamente el rendimiento general. La escalabilidad o crecimiento vertical se refiere a la flexibilidad de pasar a servidores con mejores prestaciones, o a un tipo diferente de arquitectura sin afectar severamente a los clientes (25).

La tecnología cliente/servidor realiza tres funciones básicas:

Función de presentación: Interactúa e intercambia información con el usuario, la envía para su procesamiento, permite el acceso a la aplicación, recibe los resultados de retorno y los presenta al usuario.

Función de aplicación: Procesa los datos y entrega los resultados a solicitud de la aplicación para que el usuario pueda visualizarlos.

Función de datos: recupera y asegura la integridad de los datos, de su almacenamiento y su administración.

Estas tres funciones posibilitan la distribución y buen funcionamiento de las aplicaciones a lo largo de la Red.

Los modos más comunes de usar esta arquitectura en aplicaciones web son (24):

- a. Un servidor lo hace todo: Un sólo computador realiza el servicio de **HTTP**, la negociación con el cliente, la función de datos y accede la base de datos. **ASP** (*Active Server Pages*) y **PHP** (*Personal Home Page, Hypertext Pre-processor*), son tecnologías que emplean este tipo arquitectura, ver **figura 5**.

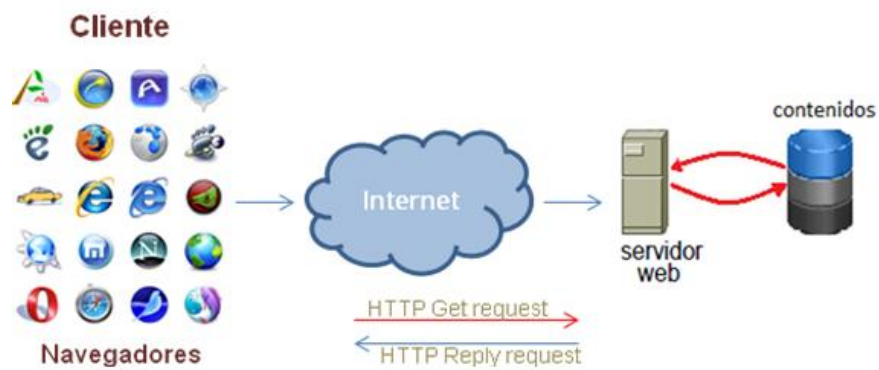


Figura 5. Arquitectura todo en un servidor

- b. Un servidor lo hace todo pero tiene un servicio de aplicaciones: en la arquitectura anterior se separa la negociación con el cliente, del servicio **HTTP** y se incluye el servicio de aplicaciones para administrar los procesos que la negociación de los datos. **JSP** (*JavaServer Pages*) es la tecnología que emplea este tipo de arquitectura, ver figura 6.

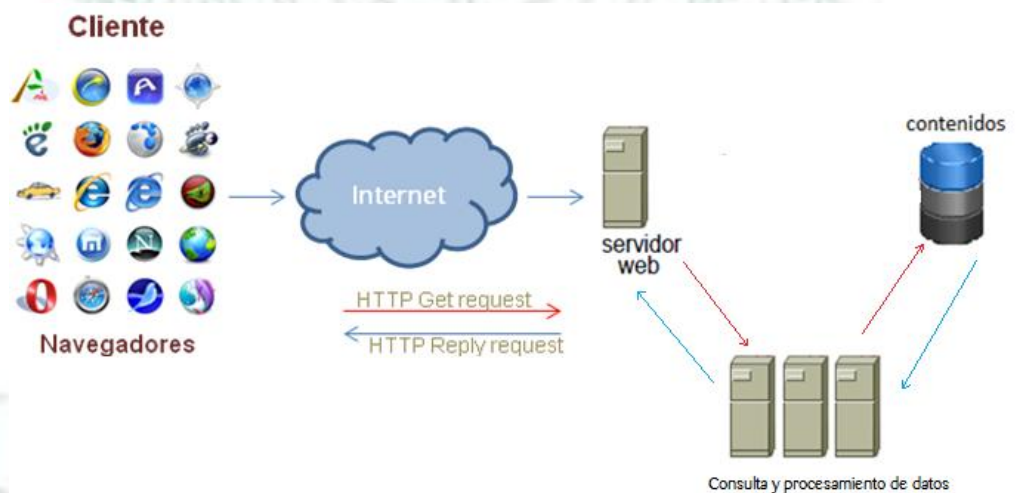


Figura 6. Un sólo servidor con servicio de aplicaciones

- c. Servidor de datos por separado pero con acceso al servicio de aplicaciones: De la arquitectura anterior, se separa la función de datos y los datos a un servidor de bases de datos aparte pero que tenga acceso al servicio de aplicaciones. **JSP** es la tecnología que emplea esta arquitectura, ver figura 7.

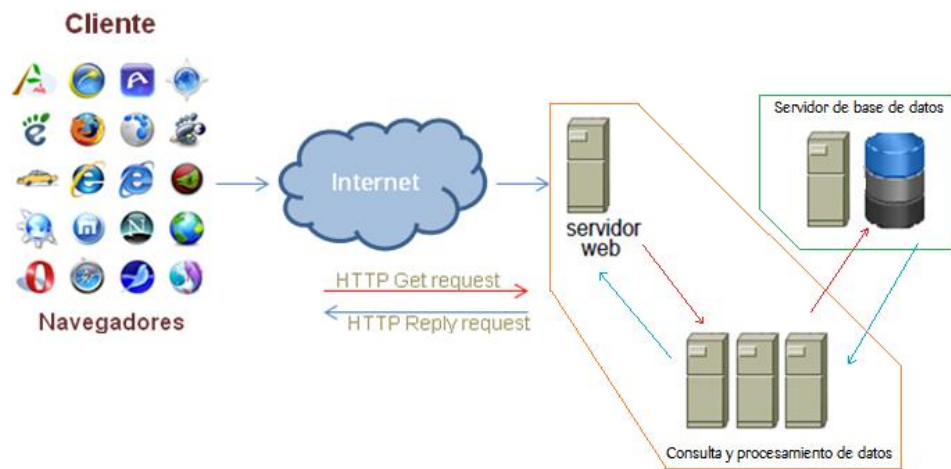


Figura 7. Servidor de base de datos por aparte

- d. El servidor web y el servidor de datos están separados: Se desprende de la arquitectura en el apartado a, pero se separa la función de datos en un servidor y la administración de los datos, desde un servidor de bases de datos aparte. No se necesita servidor de aplicaciones. ASP y PHP, son tecnologías que emplean este tipo arquitectura, ver figura 8.

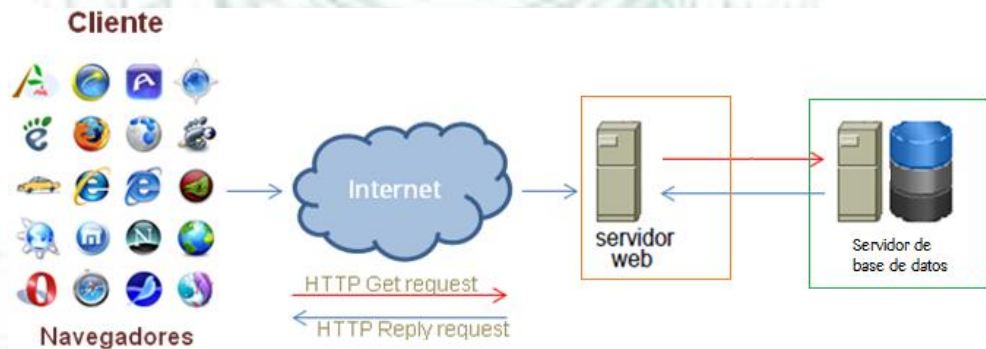


Figura 8. Sólo dos servidores

- e. Todas las funciones por separado: las tres funciones básicas en el servidor web, la negociación de los datos, la función de datos y el servicio de aplicaciones se separan en tres servidores aparte. JSP utiliza esta arquitectura, ver figura 9.

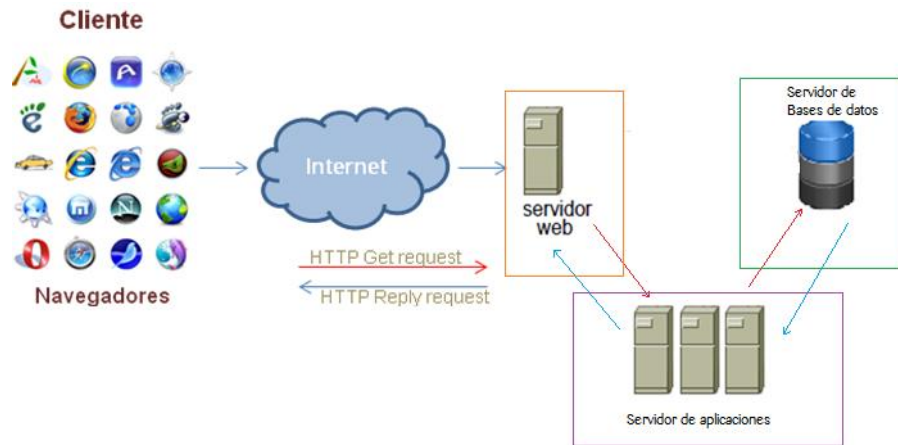


Figura 9. Tres servidores por separado

Al separar el servicio de **HTTP**, la negociación de los datos y función de datos en servidores aparte se obtiene un sistema escalable con mayor rendimiento, una configuración de requisitos flexible dependiendo de los servicios. Por ejemplo, el servicio **HTTP** necesita un computador con la buena conexión a Internet, que sea rápido pero puede tener poco almacenamiento. Sin embargo, un servidor de bases de datos necesita un computador con un disco duro rápido de alta capacidad de almacenamiento y con mucha memoria **RAM** (*Random Access Memory*) para mantener todos los datos durante el proceso de administración. Con esa topología también aumenta la seguridad de los datos, porque sólo un computador estaría conectado directamente a Internet y los demás computadores se comunican internamente (24).

2.1.1.2 El cliente web

El cliente web es una aplicación con la que el usuario interactúa para solicitar a un servidor web la entrega de recursos o información solicitada mediante **HTTP**, **FTP**, **SMTP** (*Simple Mail Transfer Protocol*), **POP** (*Post Office Protocol*) y **NNTP** (*Network News Transport Protocol*). El cliente web interpreta las páginas **HTML** (*HyperText Markup Language*) y los diferentes objetos que contienen como imágenes, videos, sonidos, etc. Los clientes web están soportados por tecnologías como: **HTML**, **Scripts**, **CSS** (*Cascading Style Sheets*), **plug-ins** (26).

Un cliente web es un navegador, un programa que puede interpretar y mostrar una página Web y capturar los clics que se hacen en diversos objetos de la página desplegada. Cuando se selecciona uno de estos objetos, el navegador sigue el enlace o hipervínculo y realiza la petición al servidor del objeto o página seleccionada. Un hipervínculo señala y nombra cualquier página u objeto que se encuentre en la Web (23).

Se usan **URLs** (*Localizadores Uniformes de Recursos*) para nombrar las páginas y los objetos, estas se divide tres secciones: el nombre del protocolo que generalmente es **HTTP** en el caso del servicio web, el nombre de dominio **DNS** (*Domain Name System*) de la máquina donde se está alojada la página y el nombre del archivo donde se está ubicada la página. Por ejemplo para la página <http://www.ideasweb.com.co/index.html> se tiene lo siguiente:

El nombre de protocolo: http://

El nombre de dominio: www.ideasweb.com.co

El nombre de archivo: /index.html

Cuando un usuario realiza un clic en un enlace o hipervínculo, o digita una dirección en la barra de búsqueda del navegador, este ejecuta una serie de pasos para obtener la página solicitada. Por ejemplo si seleccionamos <http://www.ideasweb.com.co/index.html> ocurren los siguientes pasos:

- a. El usuario escribe en el navegador el **URL** requerido
- b. El navegador solicita al **DNS** el número o dirección **IP** de www.ideasweb.com.co
- c. El **DNS** responde con el número 173.201.93.1 que es una dirección
- d. El navegador hace una conexión **TCP** con el puerto 80 en la dirección 173.201.93.1
- e. Envía un mensaje solicitando el archivo /index.html
- f. Se libera la conexión **TCP**

- g. El navegador interpreta todo el texto que se encuentra en el archivo /index.html
- h. El navegador obtiene y muestra todos los objetos enlazados por el archivo

Un navegador debe entender el formato en que se encuentra escrita la página para poder mostrarla, por esto las páginas se escriben en lenguaje **HTML** que es un estándar. El navegador puede pensarse sencillamente como un intérprete **HTML** (23).

Una página Web puede componerse de texto ordinario, texto con hipervínculos, videos, imágenes, script, etc. Sin embargo no todas las páginas pueden ser interpretadas con **HTML**. Una página puede ser en un documento con formato **PDF** (*Portable Document Format*), una imagen con formato **GIF** (*Graphics Interchange Format*) o **JPEG** (*Joint Photographic Experts Group*), música con formato **MP3**, videos con formato **MP4** o **MPEG** (*Moving Picture Experts Group*), o cualquiera otro tipo de archivo existente. El navegador no siempre reconoce todos los tipos existentes de archivos, por lo que el servidor debe entregar información sobre el tipo y subtipo **MIME** (*Multipurpose Internet Mail Extensions*) que contiene la página según las recomendaciones hechas en RFC 2045, ver **tabla 2**. Cuando el navegador es capaz de interpretar el tipo **MIME** puede desplegar los contenidos, cuando no puede, busca un plug-ins o extensión que le ayude interpretar el contenido que quiere mostrar (24).

Los plug-ins o extensiones son códigos que el navegador instala como en sí mismo, cada navegador tiene forma propia para instalar estas extensiones. En general lo que todos hacen en común es: instalar la extensión, asignarle memoria y liberarla, registrar la extensión en su directorio de ejecución y al final registrar el tipo **MIME** asociado en la extensión.

Esto puede derivar en un problema de seguridad en los navegadores, pues una extensión puede ser un archivo ejecutable con extensión .exe y el navegador lo instala como un archivo **MIME** necesario, lo ejecuta dándole permisos al programa, este puede ser altamente malicioso y estar diseñado obtener información del usuario. Muchos sitios que ofrecen información gratuita utilizan esta técnica (24).

Tabla 2. RFC 2045, Algunos tipos y subtipos MIME⁶

Tipo	Subtipo	Descripción
Texto	Plano	Texto sin formato
	Enriquecido	Texto con comandos de formato sencillo
Imagen	Gif	Imagen fija en formato GIF
	Jpeg	Imagen fija en formato JPEG
Audio	Básico	Sonido
Video	Mpeg	Película en formato MPEG
Aplicación	Octet-stream	Secuencia de bytes no interpretada
	Postscript	Documento imprimible en PostScript
Mensaje	Rfc822	Mensaje MIME RFC 822
	Parcial	Mensaje dividido para su transmisión
	Externo	El mensaje mismo debe obtenerse de la red
Multipartes	Mezclado	Partes independientes en el orden especificado
	Alternativa	Mismo mensaje en diferentes formatos
	Paralelo	Las partes deben verse en forma simultánea
	Compendio	Cada parte es un mensaje RFC 822 completo

También se puede entender los clientes como computadores de baja capacidad y clasificar en dos grupos, los independientes y los dependientes. Los primeros son computadores equipados con todo el hardware o software que les permite funcionar de manera independiente. Los segundos son computadores que necesitan un servidor para poder funcionar y sólo funcionan de manera dependiente (23).

Cuando el cliente hace una solicitud **HTTP** se produce en ocasiones una sobrecarga debido a que tiene que enviar toda la información asociada a ese dominio y la ruta juntos. Como la mayoría de los suscriptores tienen conexiones de ancho de banda asimétricas para bajar y subir información, con una relación de 1:4 a 1:20, deriva en que una solicitud de 500 Bytes en la cabecera **HTTP** puede demorarse lo mismo que subir

⁶ (24)

10 KB de datos de respuesta **HTTP** en la descarga. Esto ocurre porque los encabezados de solicitud **HTTP** se envían sin compresión. O sea, cuando se hacen solicitudes de objetos pequeños como imágenes de 10 kB, son los datos del encabezado del paquete los que utilizan la mayor cantidad de tiempo de respuesta (27).

Otro dato interesante es que la latencia se puede incrementar al comienzo de una nueva sesión del navegador. Esto ocurre porque **TCP** emplea el "comienzo lento" para nuevas conexiones y así minimizar la congestión de la red, lo que limita la cantidad de datos enviados en el inicio de la conexión antes de recibir la confirmación del destinatario. Si el que inicia la conexión sobrepasa la cantidad límite de datos a través de una nueva conexión, incurre **RTT** (*Round-Trip delay Time*) adicionales (26).

Se espera "idealmente" que una solicitud **HTTP** no exceda más de un paquete. El protocolo de Ethernet limita los paquetes de aproximadamente 1500 bytes, así que si se puede limitar las solicitudes a menos de 1500 bytes, se evita la sobrecarga por solicitud (28).

2.1.1.3 El Servidor web

Un servidor se entiende generalmente como un computador potente que gestiona archivos (servidor de ficheros), impresoras (servidor de impresoras), datos (servidor de base de datos), tráfico de red (servidor de red), páginas web (servidor web) y aplicaciones (servidor de aplicaciones) (25).

Pero cuando se habla por ejemplo de servidor web, se hace referencia un programa que está dedicado a atender las solicitudes de conexión por medio del protocolo **HTTP**, que hacen los clientes web (navegadores).

Algunas de las aplicaciones que administra un servidor web son:

Páginas estáticas o documentos **HTML** cuyo contenido no cambia.

Archivos multimedia de video, música, animaciones, etc., que se usan en de las páginas para ser visualizados directamente o descargados.

Extensiones, scripts o programas que ejecuta el servidor web cuando el cliente solicita algún objeto de la página.

Desde el punto de vista del servidor, los pasos generales que realiza el servidor cuando recibe una solicitud por parte del cliente son:

- a. Acepta la conexión **TCP** con el cliente (navegador)
- b. Recibe el nombre del archivo que solicita el cliente
- c. Obtiene el archivo del disco
- d. Envía el archivo solicitado al cliente
- e. Libera la conexión **TCP**

Sin embargo este sistema presenta un problema en su diseño y es que cada solicitud obliga al servidor a realizar un acceso al disco para obtener el archivo, por tanto un servidor web sólo puede atender tantas solicitudes por segundo como accesos al disco. Si por ejemplo el promedio acceso a un disco duro es de 5 mseg, el servidor sólo podría atender en promedio 200 solicitudes/seg (29). Se han ideado muchas soluciones a este inconveniente, las más comunes son el front-end y la memoria caché, que permiten realizar tareas multiprocesos a los servidores, pero incrementa los pasos que debe realizar el servidor para resolver una solicitud, por ejemplo los siguientes (23):

- a. Acepta la conexión TCP con el cliente (navegador)
- b. Recibe el nombre de la página Web solicitada
- c. Autentica al cliente
- d. Controla el acceso en el cliente
- e. Controla el acceso de la página web
- f. Verifica el espacio en caché disponible
- g. Accede al disco duro y obtiene la página solicitada
- h. Resuelve el tipo MIME que se incluirá en la respuesta

- i. Responde la solicitud del cliente
- j. Guarda la entrada en el registro del servidor
- k. Libera la conexión TCP

2.1.1.4 Proceso de transferencia de una página web entre cliente y servidor

Ahora se puede explicar lo que ocurre cuando un usuario cualquiera digita por ejemplo una URL <http://www.ideasweb.com.co/index.html>, o da clic a un enlace dentro de la página, el navegador captura el evento realizado por el usuario e interpreta el tipo de protocolo a usar *http://*, y el nombre DNS a buscar www.ideasweb.com.co, luego el navegador se comunica con la dirección IP del servidor y establece una conexión TCP a través del puerto 80. Establecida la conexión, envía el resto de la URL */index.html*, que es el nombre del archivo ubicado en una carpeta dentro del servidor. Éste responde la solicitud enviando el archivo, el navegador recibe el archivo, lo interpreta y lo presenta al usuario.

Recordando los pasos que realiza el cliente y los pasos que realiza el servidor, el proceso completo desde que ocurre la solicitud de una página por parte del usuario, hasta que es mostrada por el cliente web en el formato apropiado se puede describir de la siguiente manera:

- a. El usuario le dice al el cliente web la dirección de la página que quiere consultar, sea escribiendo la dirección en el navegador o pulsando un enlace
- b. El cliente web establece una conexión TCP con el servidor web a través del puerto 80
- c. El cliente web solicita la página o el objeto deseado
- d. Si existe la página u objeto solicitado, el servidor lo envía sino, responde con un código de error
- e. El cliente web realiza la interpretación de los códigos HTML si se trata de una página web sino, de los tipos y subtipos MIME. En caso de existir enlaces a objetos que deben mostrarse como sonidos, animaciones, imágenes etc, el

cliente web solicita automáticamente al servidor web dichos objetos y abre una conexión **TCP** con cada uno de ellos

- f. Se libera la conexión entre el cliente y el servidor
- g. El cliente muestra la página al usuario

Una conexión se mantiene abierta mientras se transmite la página solicitada, luego se cierra con el fin de no utilizar innecesariamente el ancho de banda disponible en la red mientras el usuario lee la página.

Cuando el usuario solicita una nueva página, a través de un enlace por ejemplo, se establece una nueva conexión **TCP**, sea para recibir otra página u objeto. De tal manera que el usuario percibe que disfrutando de una conexión permanente cuando en realidad no es así.

Es importante resaltar que para cada objeto transferido a través de la red se realiza una conexión **TCP** independiente. Por ejemplo, un el cliente web solicita la descarga de una página sencilla que sólo contiene una imagen y una animación, entonces para este caso, se establecen tres conexiones: una para el documento **HTML**, una para la imagen y otra para la animación.

2.1.2 El modelo OSI y le modelo de Internet

2.1.2.1 El modelo OSI

La **ISO** (*International Organization for Standardization*) desarrolló el Modelo de Capas **OSI** (*Open System Interconnection*), ver **figura 10**, para facilitar la comprensión de las partes que componen un sistema de comunicación y ayudar en la planeación y estandarización de las familias de protocolos. Con el fin de mantener la organización por capas los protocolos deben desarrollarse dependiendo de la capa en que se encuentren.



Figura 10. Capas del modelo OSI

Cuando una computadora intenta comunicarse con otra por medio de la Red, los protocolos de comunicación se separan dependiendo de la capa que le corresponda. Unas capas sólo puede comunicarse con la capa inmediatamente siguiente. Una forma de visualizar lo que sucede puede ser la siguiente: cuando un mensaje sale de una de las computadoras desciende por los distintos niveles de capas y cuando entra a una computadora es porque ha ascendido por los distintos niveles de capas.

Los paquetes son identificados en cada capa debido a la cabecera que cada una de ellas le agrega. Esto sirve para que la transmisión pueda ser resuelta a lo largo de la Red. De tal manera que el paquete de información original enviado por el usuario, se alarga con las cabeceras añadidas, como lo muestra la **figura 11**.



Figura 11. Cabeceras OSI anidadas

Las capas del modelo OSI cumplen la siguiente función:

Capa 1 - Física: En esta capa ocurre la transmisión de bits de datos (unos y ceros) a través de los distintos circuitos de comunicaciones. Aquí se establecen las reglas para transmitir los datos, como voltaje, el tipo de los cables, la tasa de bits enviados por segundo y conectores usados.

Capa 2 – Enlace de datos: Administra los datos provenientes de la capa 1 y corrige los errores de transmisión para las capas superiores. Crear y reconocer fronteras de mensaje marcando el punto donde comienza un mensaje y el punto donde termina. Resuelve los problemas ocasionados por mensajes perdidos, dañados o duplicados, y así protege las capas siguientes de ese tipo de errores en la transmisión. Por tanto esta capa se encarga de la detección de errores, su corrección y de retransmitir la información, también administra el tiempo de transmisión de las computadoras conectadas a al red, para que dos de ellas no lo hagan al mismo tiempo.

Capa 3 – de red: Esta encargada de enrutar los datos, decide cual es la siguiente computadora a la que debe enviar el mensaje, seleccionando la mejor ruta para que este llegue a su destino final.

Capa 4 – de transporte: Se encarga de establecer, mantener y terminar las conexiones de extremo a extremo y de los procesos que permiten entrar y salir de la red. Descompone si es necesario los grandes paquetes de datos en paquetes más pequeños, y se verifica que los paquetes sean recibidos, se encarga del control de flujos para no sobrecargar una computadora con mensajes.

Capa 5 – de sesión: Se encarga de iniciar, mantener y terminar las sesiones entre usuarios finales. Administrar, estructura y hace los arreglos necesarios para ofrecer los servicios requeridos en la sesión, como la transferencia de archivos, verificación de seguridad.

Capa 6 – de presentación: Formatea los datos para su presentación al usuario. Administra las diferentes interfaces físicas en diferentes computadoras. Exhibe, formatea y edita los datos de entrada y salida de los usuarios. Puede realizar compresión de datos, puede traducir los diferentes formatos de datos para los diferentes formatos de pantalla.

Capa 7- de aplicación: Permite el acceso del usuario final a la red. Se encarga de suministrar los datos, para el funcionamiento correcto de los servicios, a los programas de aplicación. Realiza el monitoreo de la Red y determina la acción que se debe tomar al recibir un mensaje.

2.1.2.2 EL modelo de Internet

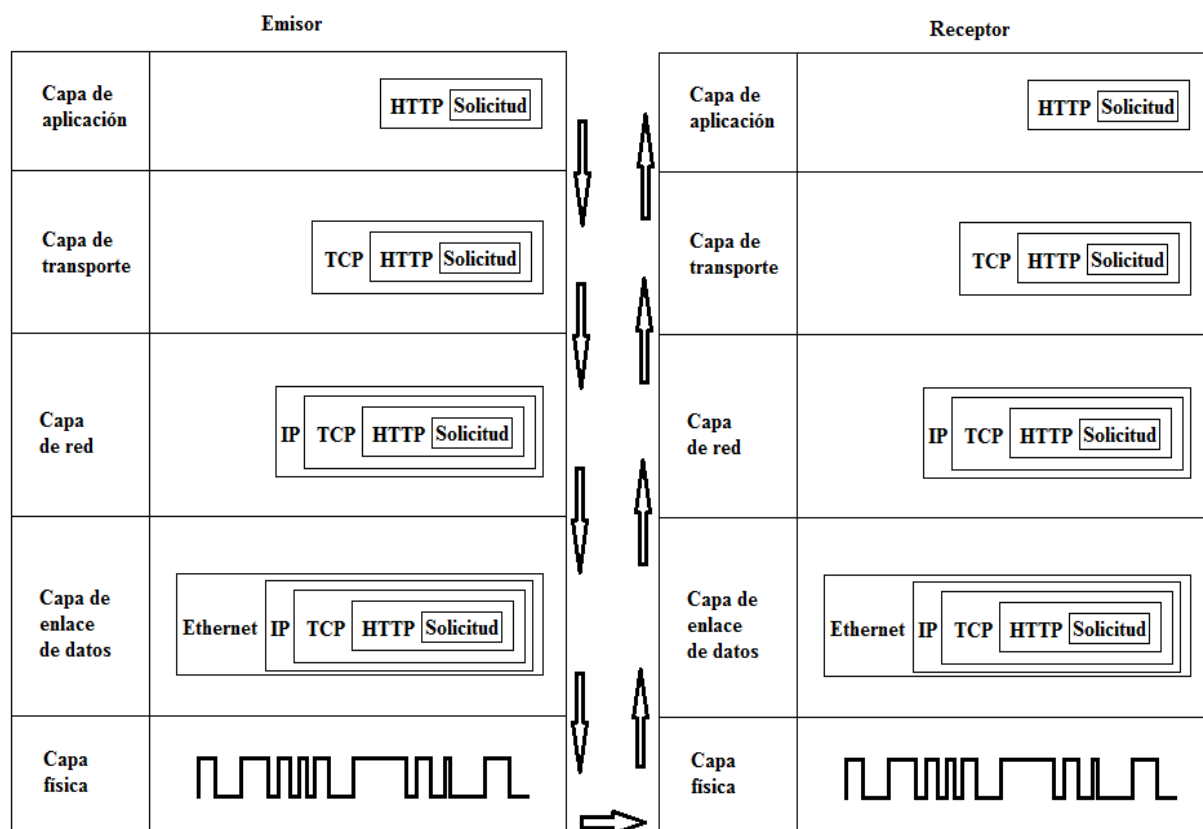


Figura 12. Capas y cabeceras del modelo de Internet⁷

Aunque el estándar oficial es el modelo **OSI**, el estándar que se está imponiendo es el modelo de Internet, más sencillo, sólo de cinco capas. El inconveniente es este modelo nunca ha sido definido formalmente y debe ser interpretado con la ayuda de varios estándares. El modelo de internet a diferencia con el modelo **OSI**, combina las tres últimas capas en una sola, la capa de aplicación. Como este trabajo se centra en la descarga de páginas web se hará referencia al modelo de Internet y no al modelo **OSI**.

⁷ (23)

Las capas en el modelo de Internet cumplen con las siguientes funciones:

Capa 1 – física: Conecta físicamente al emisor y el receptor. Se transmiten las señales eléctricas a través del circuito. En esta capa se encuentran los computadores, enrutadores, módems, los cables, satélites, antenas, etc.

Capa 2 – de enlace de datos: Se encarga de transferir el mensaje desde un dispositivo al siguiente en la ruta de la red desde el emisor hasta el receptor. La capa realiza las mismas funciones tanto en el modelo de internet como en el modelo **OSI**. Controla la capa física y da los permisos para la transmisión de los mensajes sobre los medios, detecta los errores durante la transmisión y los corrige.

Capa 3 – de red: Se ocupa de transferir los de mensaje de un extremo a otro de la red, entre el emisor y el receptor. Realiza las mismas funciones que la capa de red en el modelo **OSI**. Realiza el enrutamiento, encuentra la dirección del dispositivo siguiente si es no está en su tabla de enrutamiento y selecciona el siguiente dispositivo al que se debe enviar los mensaje.

Capa 4 – de transporte: Realiza funciones parecidas a la capa de transporte del modelo **OSI**. Realiza tres funciones. Establece la conexión entre el emisor y el receptor cuando se necesita, enlaza el software de la capa de aplicación con la red, descompone los mensajes largos en mensajes más pequeños para su transmisión y de alguna manera facilita el proceso de traducción de las direcciones utilizadas por la capa aplicación (21) (ej: www.ideasweb.com.co) para que puedan convertirse en direcciones numéricas utilizadas en las capas menores (ej: 173.201.93.1).

Capa 5 – de aplicación: Incluye las capas de sesión, presentación y aplicación en el modelo **OSI**. Se considera como el software de aplicación utilizado por el usuario de la red. Por medio de esta capa el usuario accede a la red y define los mensajes a enviar.

Ahora se hará un ejemplo sencillo de la descarga de una página web utilizando el modelo de Internet, la **figura 12** servirá de referencia para el ejemplo:

El cliente y el servidor se encuentran en la misma red, con el fin de hacer un ejemplo simple e ilustrativo.

Un usuario solicita la página web, por ejemplo www.ideasweb.com.co, a través de un navegador web (el cliente) y el computador donde se encuentra tiene la siguiente dirección 192.168.2.130. Se supone que el usuario visita frecuentemente esa página y por tanto el computador ya conoce la dirección de la capa de red y de la capa de enlace de datos del servidor Web, por esto, el software de la capa de aplicación (el navegador) conoce la dirección IP del servidor y utiliza directamente su dirección IP y no la dirección de la capa de aplicación.

El navegador entrega un paquete **HTTP**, que contiene la solicitud, a la capa de transporte solicitando una página de la dirección **IP** 192.168.3.223. La capa de transporte, en este caso **TCP**, toma el paquete **HTTP** y agrega un encabezado **TCP**, luego pasa el paquete a la capa de red, en este caso **IP**. La capa de red compara la dirección de destino con la máscara de subred 255.255.255.0 y encuentra que tanto el cliente como el servidor se encuentran en la misma subred. La capa de red va a su tabla de direcciones y busca la dirección de la capa de enlace de datos, F0-AC-23-30-BA-AC, Luego agrega un encabezado **IP** y entrega el paquete a la capa de enlace de datos, con la dirección Ethernet de destino. La capa de enlace de datos añade una trama Ethernet y lo transmite a través de la capa física hacia el servidor Web. Hasta ahora el paquete con la solicitud de la página ha descendido por las distintas capas del modelo de Internet, y comienza su ascenso hacia el servidor.

La capa de enlace de datos en el lado del servidor web verifica errores y regresa una confirmación, antes de pasar el paquete **HTTP** con los encabezados **TCP** e **IP** a la capa de red. Cuando la capa de red recibe el paquete lo procesa con el encabezado **IP**, verifica la dirección de destino y lo pasa a la capa de transporte (**TCP**), que procesa el paquete con el encabezado **TCP**, se da cuenta que es un paquete **HTTP** y lo pasa al software del servidor web.

Ahora de regreso, la respuesta del servidor web. El servidor web busca la página solicitada en el disco del computador, cuando la encuentra añade una cabecera **HTTP** y pasa el mensaje a la de capa de transporte. La capa de transporte divide la página web en paquetes más pequeños, con 1500 Bytes de longitud o menos, le agrega a cada uno un encabezado **TCP** y una numeración para indicar el orden de los paquetes. Cada uno de

ellos pasa a la capa de red, esta le agrega un encabezado **IP** donde se especifica la dirección **IP** del cliente solicitante, luego pasa el mensaje a la capa de enlace de datos junto con la dirección Ethernet del cliente y transmite los paquetes en el orden indicado por la capa de red.

La capa de enlace de datos del cliente recibe los paquetes, verifica los errores y confirma cada paquete, y los pasa a la capa de red. La capa de red verifica el destino y los pasa a la capa de transporte que ordena los datos y pasa a cada uno por turno al navegador web para que muestre la página en la pantalla.

2.1.2.3 Diferencias entre el modelo OSI y el modelo de Internet

Entre el modelo de **OSI** y el modelo de Internet existen diferencias básicas y por tanto sólo pueden compararse funcionalmente, las dos diferencias más importantes son:

1. El modelo **OSI** intenta ser más modular, por tanto separa las funciones en muchas más capas mientras el modelo de Internet agrupa más funciones en una misma capa. Por ejemplo el protocolo **TCP/IP**, en el modelo de Internet puede ser usado otros protocolos en la misma capa, mientras que en el modelo **OSI** se necesitarán dos capas más en la misma circunstancias (26).
2. En cuanto a viabilidad y eficiencia, las normas impuestas por el modelo **OSI** tienden a ser prescriptivas, mientras que el modelo de Internet y su protocolo principal el **TCP/IP** tienden a ser descriptivos, dando mayor libertad a los diseñadores e implementadores. Como la mayoría de los protocolos de Internet primero se han desarrollado y luego descrito, permiten explotar características únicas de los sistemas lo que aumenta la eficacia (27).

En la **tabla 3** se resumen las diferencias entre ambos modelos. Las columnas de estándares comunes y grupos de capas, están relacionadas al modelo de Internet, que es el modelo que se ha usado como base para toda la interpretación de los parámetros del modelo **CASUAL**.

Tabla 3. Comparativa entre el modelo OSI y el de Internet⁸

Modelo OSI	Modelo de Internet	Estándares comunes	Grupo de Capas
7. Capa de aplicación	5. Capa de aplicación	HTTP, HTML (Web), MPEG, H.323 (Audio/Video), IMAP ⁹ , POP ¹⁰ (e_mail)	Capa de aplicación
6. Capa de presentación			
5. Capa de sesión			
4. Capa de transporte	4. Capa de transporte	TCP (Internet)	Capa de trabajo ¹¹
3. Capa de red	3. Capa de red	IP (Internet)	
2. Capa de enlace de datos	2. Capa de enlace de datos	Ethernet LAN ¹²	Capa de Hardware
1. Capa física	1. Capa física	UTP ¹³ , coaxial, FTTx, etc	

2.1.2.4 Proceso de solicitud de una página a la luz de la arquitectura cliente/servidor y el modelo de Internet

Ahora que se conoce la arquitectura cliente/servidor, el modelo de Internet, y los distintos pasos del proceso al solicitar una página desde el punto de vista de cada uno de ellos, se intentará explicar el proceso de descarga de una página, incluyendo todos los conceptos vistos.

Primero, el usuario hace una solicitud en la capa de aplicación, escribiendo el nombre de la página que desea en la barra de búsqueda de un navegador web, en nuestro ejemplo www.ideasweb.com.co, o haciendo clic en un vínculo de alguna página en la que se encuentre. La capa de aplicación le añade una cabecera **HTTP** a la solicitud, con la que

⁸ (23)

⁹ **IMAP** (*Internet Message Access Protocol*)

¹⁰ **POP** (*Post Office Protocol*)

¹¹ Se refiere a que estas dos capas, son las que realizan el trabajo de control y enrutamiento, realizando la mayor parte del esfuerzo en comparación con las otras capas

¹² **LAN** (*Local Area Network*)

¹³ **UTP** (*Unshielded Twisted Pair*)

especifica el protocolo y crea así un paquete de solicitud **HTTP**. Se puede considerar al paquete de solicitud **HTTP** como un sobre que contiene el mensaje del usuario.

El navegador web coloca la información necesaria en el paquete: tipo de navegador, sistema operativo del cliente, **MIMEs** soportados, versión del protocolo **HTTP** a utilizar, y pasa el paquete a la capa de transporte. La capa de transporte utiliza el protocolo **TCP** que descompone los paquetes grandes en paquetes más pequeños, y se conecta con el servidor para transferir el grupo de paquetes. Pero en este caso el mensaje es pequeño y no necesita descomponerlo. **TCP** en su encabezado especificará que el puerto de destino es el 80.

En el caso que la capa de aplicación no conozca la dirección numérica del servidor web donde está alojada la página, el cliente hará uso del protocolo de nivel de aplicación **DNS** para ayudar a la capa de aplicación a traducir el nombre de dominio www.ideasweb.com.co a su dirección numérica 173.201.93.1. Para la explicación es suficiente con suponer que se conoce la dirección numérica.

Siguiendo la idea, la capa de transporte coloca el paquete de solicitud **HTTP** dentro de un paquete **TCP**, como un sobre dentro de otro sobre, y coloca la información necesaria como el número de paquetes que envía el orden de secuencia, así como los puertos origen y destino que identificarán la conexión. Luego toda esta información es pasada a la capa de red.

La capa de red utiliza el protocolo **IP** y selecciona la siguiente parada en la ruta que debe seguir el mensaje a través de la red. Coloca el paquete **TCP** dentro de otro sobre o paquete **IP** le añade la información necesaria: IP origen o dirección IP del Cliente (*.*) e IP destino o dirección **IP** del Servidor Web (173.201.93.1) y lo pasa a la capa de enlace datos.

Si se está accediendo a Internet a través de una **LAN**, la capa de enlace de datos hace uso del protocolo Ethernet, le coloca al mensaje marcas de inicio y fin, hace la verificación de errores y añade esa información, coloca el paquete **IP** dentro de un paquete Ethernet con la información necesaria para su transmisión: dirección MAC del

equipo cliente y dirección MAC del router que le da acceso a Internet al cliente; para así administrar la capa física en la transmisión del paquete Ethernet.

Para este caso, la capa física es el cable que conecta la computadora con el resto de la red, y la que convierte el paquete en una serie de pulsos eléctricos que viajan a través del cable.

Cuando el servidor recibe el mensaje, el proceso se repite de forma inversa. La capa física del servidor web convierte los pulsos eléctricos en unos y ceros para que la computadora pueda entenderlos y entrega el mensaje a la capa de enlace de datos. Esta última revisa las marcas de inicio y fin en el paquete Ethernet, hace la verificación de errores, y si encuentra alguno, solicita el reenvío del mensaje, pero si el mensaje no tiene errores, le quita el encabezado Ethernet y entrega el paquete a la capa de red, que revisa la dirección de destino, retira el encabezado **IP** y entrega el paquete a la capa de transporte. Esta capa procesa el mensaje, retira el encabezado **TCP** y entrega el paquete a la capa de aplicación. La capa de aplicación recibe el paquete con la información solicitada y determina cuál es el recurso o fichero solicitado por el cliente en su petición **HTTP** para buscarla en el disco.

2.1.3 Internet y TCP/IP

Internet se puede resumir en dos elementos, su forma física y el software, en su forma física un grupo de redes interconectadas entre sí, como (30):

- a. Redes troncales: Son redes de gran tamaño usada para interconectar otras redes. La conexión entre continentes y países se hace por medio de estas redes
- b. Redes regionales: Son las redes que conectan a otras más pequeñas y que se encuentran en una misma área geográfica
- c. Redes comerciales: Son las que permiten el acceso a redes troncales, por lo general pertenecen a organizaciones comerciales y pueden ser utilizadas sólo para uso interno
- d. Redes locales: Se utilizan para conectar dispositivos que están separados por unos pocos metros, por ejemplo los computadores de una oficina

Estas redes utilizan para su conexión una variedad de dispositivos como enrutadores, conmutadores, concentradores, que permiten la articulación de Internet desde los grandes centros de distribución de contenidos y los usuarios finales.

En la parte de software, o el corazón de Internet, está constituido por gran variedad de protocolos, que permiten como ya hemos visto, el intercambio de información. El protocolo por el que se reconoce Internet es el protocolo **TCP/IP**.

2.1.3.1 Protocolos de Internet

Muchos de los servicios que ofrece Internet hace parte del protocolo **TCP/IP**, que permiten el intercambio de información independientemente de los sistemas en que ésta se encuentra almacenada, lo que le da mucha versatilidad debido a la gran variedad de los sistemas informáticos. En el grupo de protocolos **TCP/IP** están varios protocolos de nivel de aplicación de acuerdo al modelo de Internet y que proporcionan una serie de servicios.

Como en un computador se puede atender distintos servicios simultáneamente, se distribuyen las solicitudes al llegar sobre distintos puertos, que corresponden al tipo de servicio que se requiera. Algunos servicios son tan comunes que ya tienen reservado en forma estándar el puerto al que deben llegar. Ver **tabla 4**.

Tabla 4. Protocolos comunes en Internet¹⁴

Nombre	Acrónimo	Puerto	Descripción
File Transfer Protocol	FTP	21	Transferencia de ficheros
Telnet		23	Conexión en modo terminal a sistemas remotos
Simple Mail Transfer Protocol	SMTP	25	Envío de correo electrónico
Domain Name Server	DNS	53	Resolución de nombres de dominio
Hyper Text Transfer Protocol	HTTP	80	Transferencia de páginas web
Post Office Protocol v3	POP3	110	Recepción de correo electrónico
Network News Transfer Protocol	NNTP	119	Acceso a foros de discusión

¹⁴ (24)

2.1.3.2 El protocolo TCP/IP

Este conjunto de protocolos está dividido en dos partes, como su nombre lo indica, el protocolo **TCP** ubicado en la capa de transporte y el protocolo **IP** ubicado en la capa de red. La parte **TCP** se encarga de descomponer los paquetes en otros más pequeños, numerarlos y organizarlos de forma correcta según su orden, así como de identificar cada conexión mediante el uso de un par de puertos, uno local y otro remoto. La parte **IP** se encarga de enrutar y direccionar los paquetes según sea su destino final. El protocolo de **TCP** es el intermediario entre la capa de red y la capa de aplicación.

Como se observa en la **figura 13**, el encabezado **TCP** tiene 192 bits de información de control. Contiene los identificadores de los puertos de origen y destino, el número de secuencia enviado, el número de secuencia recibido, el margen de datos que permite saber donde comienza el paquete, un espacio reservado que no se usa, banderas (**URG**-urgente, **ACK**-acuse de recibo, **PSH**-opera con el método PUSH, **RST**-restablecer conexión, **SYN**-establecer conexión, **FIN**-interrumpe conexión), un espacio para indicar cuantos bytes que no necesitan acuse de recibo, un apuntador urgente que indica a partir de qué número de secuencia el mensaje se torna urgente, un espacio para varias opciones y un espacio para los datos del usuario (31).

El protocolo **IP** hace parte de la capa de red. En esto momentos se encuentran activas dos formas del **IP**. La forma tradicional conocida como la versión 4 del **IP** (**IPv4**), y al igual que **TCP**, también tiene un encabezado de 192 bits. Como se observa en la **figura 13**, este encabezado contiene un espacio donde especifica la versión **IP** que usa, un espacio para la cantidad de palabras de 32 bits que componen el encabezado, un espacio para el tipo de servicio, un espacio para indicar el tamaño del paquete que no puede exceder de 65536 bytes, un espacio de identificación, un espacio para banderas, un espacio para los márgenes del fragmento, un espacio donde se indica el tiempo de vida del paquete, un espacio para el tipo de protocolo de la capa de transporte **ICMP** (*Internet Control Message Protocol*), **IGMP** (*Internet Group Management Protocol*), **TCP**, **UDP** (*User Datagram Protocol*), un espacio para la suma de comprobación, un espacio para la dirección de origen, uno para la dirección de destino, un espacio para otras opciones y los datos del usuario (32).

El IPv4 está siendo reemplazado por una nueva versión, el IPv6, que tiene un encabezado de 320 bits. Esto es necesario porque el rápido crecimiento de Internet está agotando las direcciones IP, un formato mayor permite más direcciones. Pero los cambios incluidos en el protocolo IPv6 implican mejoras en el protocolo TCP, por lo que se espera que se desarrolle una nueva versión de TCP (33).

Los datos de la solicitud del usuario son variables por tanto el espacio ocupado en este campo depende del protocolo utilizado. El TCP/IP se combina en la capa de enlace de datos con el protocolo Ethernet cuyo tamaño de paquete máximo es de 1492 Bytes, por lo que tamaño máximo de los datos de usuario en un paquete de este tipo, si se utiliza el IPv4 es de 1492 - el tamaño del encabezado TCP (24 Bytes) - el tamaño del encabezado IPv4 (24 Bytes) = 1444 Bytes. Ver **figura 13**.

Ahora se hará un pequeño acercamiento de lo que pasa con el protocolo TCP cuando se hace una solicitud, para esto es importante entender dos conceptos que además servirán luego para entender los modelos de Teletráfico:

Esquema orientado a conexión: EL protocolo TCP establece una conexión entre el emisor y el receptor por medio de un circuito virtual. Un circuito virtual es una simulación de un circuito físico, en este caso creada para hacerle parecer a la capa de aplicación que utiliza un circuito punto a punto, pero en realidad no es así. Para esto, la capa de transporte envía un paquete con la bandera SYN para establecer la conexión con el receptor. Una vez establecida la conexión, todos los paquetes del mismo mensaje viajan a través del circuito virtual y llegan al destino en el mismo orden en el que fueron enviados, esto facilita la reorganización de los paquetes en el receptor (23).

UNIVERSIDAD
DE ANTIOQUIA
1803



Paquete del protocolo de control de transmisión (TCP). ACK = confirmación; CRC = verificación de redundancia cíclica

Puerto de origen	Puerto de destino	Número de secuencia	Número de ACK	Longitud del encabezado	No utilizado	Banderas	Control de flujos	CRC-16	Apuntador urgente	Opciones	Datos del usuario
16 bits	16 bits	32 bits	32 bits	4 bits	6 bits	6 bits	16 bits	16 bits	16 bits	32 bits	varia

Paquete del protocolo IPv4

Número de versión	Longitud del encabezado	Tipo de servicio	Longitud total	Identificadores	Banderas	Desplazamiento de paquete	límite de brincos	Protocolo	CRC-16	Dirección de origen	Dirección de destino	Opciones	Datos del usuario
4 bits	4 bits	8 bits	16 bits	16 bits	3 bits	13 bits	8 bits	8 bits	16 bits	32 bits	32 bits	32 bits	varia

Paquete del Protocolo IPv6

Número de versión	Prioridad	Etiqueta de flujo	Longitud total	Siguiente encabezado	límite de brincos	Dirección de origen	Dirección de destino	Datos del usuario
16 bits	16 bits	32 bits	32 bits	4 bits	6 bits	38 bits	128 bits	varia

Figura 13. Cabeceras IPv4 y IPv6¹⁵

La técnica usada por TCP en la transmisión es conocida como ventana deslizante o ARQ (*Automatic Repeat-reQuest*) continua, porque envía todos los paquetes sin esperar la confirmación, y cuando recibe la confirmación de los paquetes que llegaron con error, los retransmite. Sólo retransmite los paquetes que llegaron con error. Cuando se termina la transmisión de paquetes con éxito, el emisor envía un paquete final con la bandera FIN y con esto se cierra la conexión, se cierra el circuito y se todo registro de éste (31).

Esquema sin conexión: A diferencia del anterior, aquí no se crea un circuito virtual, por tanto no es necesario que se establezca conexión entre el emisor y el receptor. Cada paquete viaja por separado a través de la red. El emisor envía cada paquete de forma separada y cada uno de ellos puede tomar diferentes rutas a través de la red, viajar a distintas velocidades y llegar al destino en cualquier orden. Como los paquetes pueden acumular retardos muy grandes, la capa de red en el emisor agrega una secuencia en cada paquete para que la capa de red en el receptor pueda reensamblarlos en el orden correcto y formar el mensaje antes de pasarlo a las siguientes capas (23).

TCP¹⁶/IP es muy versátil porque puede operar en los dos modos, orientado a conexión o sin conexión. Cuando opera orientado a conexión, se establece el circuito virtual con el

¹⁵ (31)

¹⁶ Aunque la capa de control de Internet en general es conocida con el nombre TCP, realmente en esta capa existen dos protocolos o modos de trabajo, de igual nivel e importancia, el protocolo TCP y el UDP

destino y el protocolo **IP** enruta todos los paquetes a través de este circuito virtual hasta su destino. Cuando opera sin conexión, el paquete **TCP** se reemplaza por un paquete del Protocolo de Datagrama de Usuario **UDP**. Este tipo de paquete es 8 bytes más pequeño que el paquete **TCP** y únicamente contiene el puerto de origen, el puerto destino, la longitud del mensaje y la suma de verificación (34).

Generalmente el modo sin conexión es usado cuando los datos de aplicación o el mensaje caben en un solo paquete¹⁷. Sin embargo, a pesar que las solicitudes **HTTP** comúnmente son muy cortas, no utilizan esquema sin conexión **UDP**, sino enrutamiento orientado a conexión **TCP**. **HTTP** siempre utiliza el **TCP** y la mayoría de los protocolos mencionados hasta ahora (**HTTP**, **SMTP**, **FTP**, **Telnet**) utilizan **TCP**. El protocolo **UDP** se utiliza mucho para mensajes de control como direccionamiento **DHCP** (*Dynamic Host Configuration Protocol*), , enrutamiento **RIP** (*Routing Information Protocol*), Protocolo de Información de Enrutamiento), y administración de redes **SNMP** (*Simple Network Management Protocol*) (34).

2.1.3.3 ¿Cómo son las conexiones TCP?

Necesariamente cuando una computadora intenta transmitir datos hacia otra computadora, debe decidir entre un servicio orientado a la conexión **TCP** o un servicio no orientado a la conexión **UDP**. Los clientes web que usan el servicio **HTTP** siempre emplean esquema orientado a conexión. Por este motivo, antes de enviar el primer paquete, la capa de transporte primero envía un paquete con una bandera **SYN** y establece una conexión. Luego de establecida la conexión, envía todos los paquetes de datos y cuando termina, envía un paquete con bandera **FIN** para cerrar la conexión (34).

El primer paquete con la bandera de sincronización **SYN** y el último paquete con la bandera de finalización **FIN**, se direccionan y se enrutan del mismo modo que cualquier otro paquete **TCP**.

En un principio cuando se desarrolló **HTTP**, los navegadores web establecían una conexión **TCP** por cada solicitud **HTTP**. Cuando realizaban la solicitud de una página por ejemplo, sucedía lo siguiente: se establecía la conexión, luego enviaba un paquete

¹⁷ Esto no es del todo cierto, depende del tipo de servicio

solicitando la página web, el servidor web luego de aceptar la conexión, enviaba todos paquetes necesarios para transmitir la página solicitada y al terminar cerraba la conexión. Ahora, si la página tenía imágenes, el navegador web establecía y cerraba una conexión diferente por cada solicitud de imagen. Con el crecimiento de Internet se detectó que ese procedimiento aumentaba el consumo de ancho de banda y consumía más tiempo del necesario. La versión 1.2 de **HTTP**, permite a los navegadores web abrir la conexión cuando envían por primera vez una solicitud **HTTP** y dejar abierta la conexión para todas las demás solicitudes derivadas en la misma conexión y hechas al mismo servidor.

2.2 La WEB

Para comprender el servicio web es inherente a entender la Web. Internet y la Web son conceptos distintos, sin embargo es muy común referirse a ellos como una misma cosa. Como ya se ha dicho anteriormente Internet es un conjunto de redes articuladas gracias a un conjunto de protocolos que permiten la administración y transferencia de información. La Web sin embargo, es el conjunto de sitios, páginas y objetos que llenan de contenido una pequeña parte de Internet. La Web necesita de algunos protocolos que ofrece Internet para poder operar. Es muy fácil que las personas confundan estos dos conceptos, porque los navegadores web, son la interfaz de acceso a la incalculable cantidad contenidos almacenados en los millones de servidores que existen en el mundo. La Web está constituida por tres elementos claves:

- Todos los contenidos de la Web están asociados a una URL como medio de localización.
- El protocolo encargado de la transferencia de los contenidos y recursos de la Web es el **HTTP**
- El lenguaje con el que se construyen los contenidos de la Web es el **HTML**

2.2.1 ¿Qué es un sitio web?

Un sitio web puede entenderse en su concepto más amplio como un conjunto de páginas web que están relacionadas entre sí por medio de enlaces, y por lo general asociadas de alguna forma a la misma **URL**.

Una página web puede entenderse como un fichero que contiene un código generalmente **HTML** que indexa o llama a unos recursos que se usan para enriquecer el contenido de la página o mejorar su aspecto, como sonidos, videos, imágenes.

Cuando un usuario digita la **URL** de una página en su navegador web, puede recibir la página de inicio del sitio o una página de bienvenida. Sin embargo ambas tienen unas características en común: son usadas para promocionar algún producto o servicio y el nombre de la compañía propietaria del sitio web, también sirven como puerta de entrada a los recursos disponibles en el sitio web, y que pueden ser privados o públicos. Estas páginas suelen ser llamativas, con nada o poco de texto, muchas imágenes, animaciones, videos e incluso sonidos.

Otro de los usos de estas páginas es que sirven para disminuir el tiempo necesario de descarga para las páginas siguientes, ya que se aprovecha el tiempo que el usuario utiliza para visualizar página inicial, para descargar en segundo plano y guardar en la memoria caché del navegador las imágenes que se usaran en las páginas posteriores.

La diferencia esencial entre la página de bienvenida y la página principal o home, es que la última funciona como índice de contenidos, donde se listan los distintos enlaces de interés que componen el sitio web y que llevan a otras páginas o sitios.

2.2.2 ¿Qué es una aplicación web?

Es un medio que permite realizar peticiones a una aplicación remota accesible por medio de Internet y que recibe una respuesta que se muestra a través de la misma aplicación. El ejemplo más claro de una aplicación web es un navegador. También se entiende como una aplicación web, las interfaces que permiten interactuar remotamente con un software instalado en un servidor, como ejemplo tenemos los buscadores.

Las aplicaciones web pertenecen a la arquitectura cliente/servidor, un cliente solicita servicios mientras el servidor espera recibir solicitudes y responderlas. Para que esto suceda se utilizan recursos que permitan programar aplicaciones web, algunas del lado del cliente como ActiveX o applets, sin embargo como ya se ha dicho los estándares de facto son **HTML** y JavaScript y otras del lado del servidor web como **ASP**, **CGI** (*Common Gateway Interface*), **PHP**, **JSP**, etc.

Una aplicación web es un tipo de aplicación cliente/servidor, donde el cliente web (navegador), el servidor web, y el protocolo (**HTTP**) mediante el cual se comunican son un estándar y no son creados por el desarrollador de la aplicación.

2.2.3 Tipos de objetos

Las páginas web pueden ser visualizadas y entendidas por el usuario gracias a los distintos tipos de objetos que se usan para su representación, los objetos más comunes son datos, texto **HTML**, Scripts, imágenes, video y audio.

Los datos: son la representación de la información en forma de caracteres de texto. No necesitan mucha capacidad de almacenamiento, y tampoco exigen altas tasas de transmisión. La transmisión de los datos depende de los protocolos del servicio implementado en el sistema de comunicación y para cada servicio presenta patrones diferentes. Por ejemplo los accesos a servidores web realizan grandes cantidades de transferencia de paquetes pequeños, mientras que la transferencia de ficheros es larga y con periodos de silencios entre ellos (35).

HTML: es un lenguaje de programación basado en etiquetas utilizado para dar forma a los contenidos y documentos que se publican en la Web. Los navegadores interpretan las etiquetas **HTML** y organizan la información para mostrarla al usuario.

En general las páginas web están escritas en el lenguaje **HTML**. Sin embargo la evolución de la Web ha permitido el desarrollo de páginas más complejas escritas en otros lenguajes como **DHTML** (*Dynamic HTML*), **CSS**, JavaScript, applets o componentes ActiveX.

Las etiquetas de **HTML** permiten modificar tipos de letras, colores, imágenes, tablas, listas de elementos, etc. Usarlo es muy sencillo, sólo debe añadirse la etiqueta la comienzo y al final del contenido que se quiera modificar.

Scripts: son lenguajes que permiten incluir programación dentro del código **HTML** de las páginas web y que se pueden ejecutar con o sin el consentimiento del usuario. En otras palabras, es un lenguaje de programación, que se usa dentro del contexto de una aplicación, con la limitante de no permitir la programación aplicaciones independientes. Estos lenguajes son interpretados, aunque en ocasiones pueden tener un proceso de pseudo-compilación usado para optimizar su ejecución. Los más usados en el desarrollo de páginas web son JavaScript, VBScript (34).

Imágenes: El tipo de imágenes que son parte de nuestro interés, son aquellas conocidas como digitales y que se pueden usar a través de medios informáticos. Este tipo de imágenes poseen distintos formatos. Lo más usados en internet son JPEG, GIF, BMP y **PNG** (*Portable Network Graphics*).

En la siguiente **tabla 5** se resumen las características principales de los formatos.

Tabla 5. Algunos formatos de imágenes y sus características¹⁸

Característica	GIF	JPEG	PNG
Colores	256 (8 bits)	16777 (24 bits)	48 bits
Transparencia	Sí	No	Sí (alfa)
Animación	Sí	No	No
Compresión	Sin pérdidas	Con pérdidas	Sin pérdidas
Dibujo	Sí	No	Sí
Fotografía	No	Sí	Sí
Gamma	No	No	Sí

- a. **GIF:** Es un formato de imagen que permite representar un máximo de 256 colores por pixel. En general el navegador descarga y muestra la imagen en el modo entrelazado, y no secuencialmente, lo que quiere decir que primero

¹⁸ (24)

muestra las filas impares de pixeles de la imagen y luego las filas pares. La ventaja de este procedimiento es que permite una vista previa de la imagen que se está descargando

- b. **JPEG**: es un formato de imagen que permite representar un máximo de $16'777.216$ colores por pixel para mostrar fotografías o ilustraciones con alta calidad. El navegador descarga este tipo de imágenes de forma progresiva, lo que quiere decir que muestra las filas de pixeles que forman la imagen una tras otra en forma secuencial
- c. **PNG**: es un formato desarrollado por la **W3C** para internet, que mejora características del formato **GIF** como la representación de 2^{48} colores por pixel y la transparencia, pero aprovecha características como la descarga entrelazada de filas de pixeles para formar la imagen en el navegador
- d. **BMP (BitMaP)**: es un formato de codificación de bits conocido como mapa de bit, fue desarrollado por *Microsoft* y utiliza distintas resoluciones como 1, 4, 8, 16, 24 ó 32 bits/*pixel*. Por su sistema de compresión **RLC (Run Length Code)** es muy útil para imágenes que presentan abundancia de un mismo color. La forma en que lo muestra el navegador es progresiva

Video: Al igual que las imágenes, es necesario aclarar que el video de nuestro interés es el que se encuentra en formato digital. Un video es una secuencias temporal de imágenes estáticas, que al mostrarlas una a una de forma consecutiva dan la sensación de movimiento. Esto es posible porque la diferencia entre dos imágenes consecutivas es mínima. Como la variabilidad es mínima entre cada imagen y los pixeles contiguos son muy parecidos entre ellos se utiliza un esquema de predicción hacia adelante y hacia atrás sacándole mayor provecho a la redundancia espacial y a la redundancia temporal. Esto es conocido como dependencia autorregresiva de fuentes de tasa variable. El formato de video más usado es el **MPEGxx**, sin embargo existen más como **VCD (Video Compact Disc)**, **DIVX**, **HDTV (High Definition TV)**, **DVD (Digital Versatile Disc)**, **XVID** entre otros (35).

Audio: del mismo modo que en los formatos de imagen y video sólo interesa el audio digital. En general las características de este tipo de audio depende de los codecs usados, de su compresión y de su transmisión, la cual a su vez definen su calidad. Los protocolos usados para el audio son **SIP** (*Session Initiation Protocol*), **IAX** (*Inter Asterisk eXchange*), **H.323**, **MGCP** (*Media Gateway Control Protocol*) y **SCCP** (*Skinny Client Control Protocol*). Y los formatos más comunes de audio son **MIDI** (*Musical Instrument Digital Interface*), **MP3** (*MPeg audio layer 3*), **WAV** (*Waverform Audio File*), **RA** (*Real Audio*). Este último formato es de particular importancia porque es usado para transferir archivos de audio en tiempo real, de tal manera que el archivo se va reproduciendo en la medida que está siendo descargado (35).

En resumen, los objetos se someten a un proceso de compresión/codificación para optimizar el ancho de banda, para estos se utilizan diferentes *codecs*: **MPEGx** para video, **T.1xx** para datos; **JPEG**, **BMP**, **GIF**, **PNG** para imágenes fijas, etc; lo que permite enviar/recibir la señal como secuencia continua de paquetes, agrupados en bloques de datos, a una tasa de datos determinada, y todo en función del *codec* empleado. A cada uno de estos paquetes se le añaden cabeceras de las distintas capas que componen el modelo de Internet y que ocupan un número variable de bytes de acuerdo al tipo de transmisión. En la **tabla 6** se puede ver el tipo de objeto, los formatos más comunes, el tamaño de los paquetes y el servicio y aplicación que los usan.

UNIVERSIDAD
DE ANTIOQUIA
1803



Tabla 6. Tipos de objetos principales: características y parámetros¹⁹

ToS	codec	algoritmo	Rd ²⁰ (kb/s)	Sp ²¹ (bits)	fm/h	servicios/aplicaciones	refs
Audio	G.711	PCM-leyA	48 - 64	6-8(β =268B RTP)	8kHz / -	Telefonía analogH.320	[113]
	G.721	ADPCM	24 - 32	3-4(β =268B RTP)	8kHz / -	Audio conferencia	[115]
	G.722	ADPCM	48 - 64	6-8(β =268B RTP)	7kHz / -	Videoconferencia	
	G.726	ADPCM	16 - 40	2-5(β =268B RTP)	N/A	Voz digital en red	
	G.727	ADPCM	16 - 40	2-5(β =268B RTP)	N/A	Voz digital en red	
	G723.1	MP-MLQ	5.30	160(β =32BRTP)	33kHz / -	Telefonía por Internet	
		ACELP	6.40	184(β =36BRTP)	30kHz / -	(MACdelay 37.5ms)	
	G.728	LD-CELP	16	10(β =368RTP)	1.6kHz / -	VoIP (Voz sobre IP)	[116]
	G.729	CSACELP	8	80(β =368RTP)	100Hz / -	VoIP redes Internet	[117]
	CELP	LPC-10	2.4 - 4.8	20 - 24(β =248RTP)	N/A	(MACdelay 22.5ms)	[118]
	GSM06.10	RPE-LTP	13.2-22.8	240- 20(β =48BRTP)	50Hz / -	Telefonía móvil GSM	[129]
	AMRxxx	AMR	4.7-12.2	95-244(β =8B RTP)	50Hz / -	Telefonía móvil	[130]
	MPEG-1	DolbyAC3	32 - 448	N/A	N/A	Música digital	[119]
MP3.Wav		24 - 128	N/A	N/A	Música por Internet		
Datos	HTTP	HTTP	12-32.	20-40-100-400k	1/200-500	accesos vía web	
	FTP	FTP	100 - 720	1-10-100-1000k	1/1-5-10k	transmisión ficheros	
	TELNET	TELNET	8-40.	10-20-40-100k	1/100-400	acceso remoto	[120]
	SMTP	SMTP	4-24.	50-100-200-400k	1/10-50k	envío avisos/mensajes	[121]
	SNMP	SNMP	12 - 200	10-20-40-100k	1/400-800	gestión remota equipo	
Imagen	JPEG	24b/píx	50 - 200	640 x 480píx	4:2:0(4Y1U1V)	Formato genérico	[122]
					4:2:2(4Y2U2V)	múltiples aplicaciones	
					4:4:4(4Y4U4V)	imágenes fijas	
	BMP	RLC	40 - 400	640 x 480píx	4-8b/píx	Imágenes windows	[123]
	GIF	LZ	10 - 200	640 x 480píx	8b/píx	Imágenes simples	[124]
PNG	HuffmanLZ	10 - 100	640 x 480píx	48b/píx	Imagen alta resolución	[125]	
Video	H.120	H.32x	1.5M	CIF 352x288	30/1000	En desuso	[131]
	H.261	k=1...30	k-64	CIF 352x288	30/1000	Teleconferencia	[132]
		348k-2M	110 - 320	QCIF 176x144	15/1000	Videófono Digit H.320	
	H.263	5-15 fps	< 64	SQCIF 192x144	30/1000	Teleconferencia	[133]
		15-25fps	20 - 200	4CIF 704x576	30/1000	Videófono anlgH.324	
		25-30fps	24.8 - 768	16CIF 1048x1152	30/1000	Telefonía GSM	
	MPEG-1	MPEGx	1.5M	QCIF 352x240	0.4-2M	Vídeo PC, videojuegos	[134]
	MPEG-2	MPEGx	4-100M	QCIF 704x576	1.5-60M	TV Digital	
	MPEG-3	MPEGx	4-9M	QCIF 704x576	N/A	HDTV alta calidad	
	MPEG-4	MPEGx	4.8-64k	QCIF 176x144	28.8-500k	Videoconferencia	
	MJPEG	MJPEG	10-20M	QCIF 352x288	N/A	Experimental	
ShowMe (SunSparc)	24b/pix	1.7M	QCIF 176x144	1-30/256c	Videoconferencia (EED < 220-300ms)	[276]	
			QCIF 640x480	1-30/256c			
			QCIF 384x288	1-15-30			

¹⁹ (35)

²⁰ Rb = tasa de bit (b/s), Rd = tasa de datos (b/s)

²¹ Sp = tamaño de paquetes (B), S = tamaño de bloque de datos (B)

3. EJE DE ACCESOS

Es importante conocer las tecnologías que soportan las redes de comunicaciones, conocidas como tecnologías de acceso al medio, para nuestro caso Internet. Estas pueden ser fijas o inalámbricas. Al final de esta sección, en la **tabla 7** se encuentra el resumen de las tecnologías más representativas, a continuación se presentan algunas de ellas.

3.1 Tecnologías de acceso fijas

- *Línea Telefónica Básica, PSTN (Public Switched Telephone Network)*. Usada para la transmisión de voz y datos aunque no con mucha velocidad pues es una línea telefónica analógica, lo que implica una limitación en la frecuencia que oscila entre los 300 a 3400Hz para los datos y con la ayuda de un modem, se conoce como banda estrecha porque ofrece una transferencia máxima de 56 kb/s a una distancia corta (35).
- *Línea de Acceso Digital, DSL (Digital Subscriber Line)*. Usada para transmitir simultáneamente sobre el mismo par de cobre, voz y datos a distintas velocidades para que no se interfieran. Para lograrlo se utilizan dos modem, uno en la residencia del usuario y otro en la empresa prestadora del servicio de Internet, cada uno de estos modem necesita un filtro que separa las dos señales. La distribución del ancho de banda es asimétrico, o sea que ofrece mayor tasa de transferencia de datos en una de las dos direcciones, generalmente desde la central hacia el suscriptor, para brindarle mayor capacidad de descarga, aproximadamente entre 256 kb/s hasta 1.5 Mb/s y en la otra dirección, desde el suscriptor hasta la central, entre 64 kb/s hasta 256 Mb/s (36).
- *Línea de Red Digital, ISDN (Integrated Services Digital Network)*. Permite comunicación punto a punto digital, a diferencia de la red telefónica analógica tradicional. Garantiza muy buena calidad, pocas interferencias y la tasa de transmisión contratada al ISP. Con esta tecnología se prestan servicios de voz, datos y multimedia. Se puede tener acceso a diferentes tipos de redes de información a través de interfaces **BRI (Basic Rate Interfaces)** y **PRI (Primary Rate Interfaces)** con tasas de transferencias de 64kb/s simétricos usuario–central y generalmente agrupados en N canales de $N * 64\text{kb/s}$ (36).

• *Línea coaxial*: Esta tecnología les permite a las operadoras de televisión por cable proporcionar acceso a Internet de banda ancha por medio de un cable módem y haciendo uso de los cables coaxiales que se usan para la señal de televisión. El cable módem permite de datos para el computador de la señal de imagen y sonido de la proporcionada para el televisor. Esta tecnología de acceso de banda ancha tiene la capacidad de permanecer siempre conectado y ofrecer altas velocidades. Las velocidades de usuario residencial oscilan alrededor de 1.5Mb/s (35).

• *Línea fibra*: Esta tecnología ha evolucionado mucho en los últimos años aminorando su costo que ha hecho posible llevarla hasta el hogar. Con este tipo de conexión aparece la integración de telecomunicaciones en distintos tipos de servicios, la telefonía, el acceso a Internet de banda ancha, televisión, VoD (*Video on Demand*), telepresencia y otros. El acrónimo general usado para referirse a esta tecnología es **FTTx** para referir los distintos accesos ofrecidos con ella, **FTTN** (*Fiber-To-The-Node*), **FTTC** (*Fiber-To-The-Curb*), **FTTB** (*Fiber-To-The-Building*), **FTTP** (*Fiber-To-The Premises*), **FTTD** (*Fiber-To-The-Desk*).

3.2 Tecnologías de acceso inalámbricas

• *Acceso por satélite*: Por este tipo de acceso se pueden transmitir señales de televisión, teléfono, y datos. La velocidad del acceso por satélite depende de la línea visual entre la antena y el satélite que está en órbita, el clima, el proveedor de servicio, etc. El rango de la tasa de transmisión típica va desde 64kb/s hasta 2Mb/s. La principal ventaja de este medio es que permite acceso desde cualquier parte del mundo, sin la necesidad de otras redes.

• *Acceso por radio fija*: Las más importantes son **WiFi** (*Wireless Fidelity*), **WiMax** (*Worldwide Interoperability for Microwave Access*), y **LMDS** (*Local Multipoint Distribution Service*). **WiFi** permite tasas de hasta 54Mb/s dependiendo de la norma **IEEE** (*Institute of Electrical and Electronics Engineers*), **WiMax** permite tasas hasta 70Mb/s dependiendo de la norma **IEEE**, **LMDS** permite tasas de hasta 42Mb/s dependiendo del número de canales y la modulación.

• *Acceso móvil*: Las tecnologías móviles han evolucionado en los últimos años y es muy común que las personas tengan dispositivos celulares. Son un recurso muy flexible. El comienzo de esta tecnología se encuentra a principios de 1990 cuando fue implantada la **GSM** (*Global System Mobile*) con tasas de transmisión de 9.6kb/s y que funcionaba de forma similar a una **PSTN** pero



en entornos móviles. Luego aparecen evoluciones **HSCSD** (*High Speed Circuit Switched Data*) que permite tasas de 28.8–43.2kb/s, **GPRS** (*General Packet Radio Service*) con tasas de 115-160kb/s, **EDGE** (*Enhanced Data Rates for GSM Evolution*) con tasas de 384–473.6kb/s, **UMTS** (*Universal Mobile Telecommunications System*) con tasas de 384kb/s–2Mb/s, **HSDPA** (*High Speed Downlink Packet Access*) con tasas de 3.6-7.2 Mb/s, **HSPA** (*High-Speed Packet Access*) con tasas de 42 Mb/s y finalmente la última tecnología móvil que se imponiendo es **LTE** (*Long Term Evolution*) con tasas de 100MB/s – 1Gb/s (35).

Tabla 7. Distintas tecnologías de acceso

Tecnología	BW (b/s)		Costo	Tipo de aplicación
	Download	Upload		
PSTN	33.6k	56k	bajo	Datos/Audio
DSL	>1M	>2-300k	bajo	Multimedia
ISDN	128k(BRI)	1.5-2M	alto	Multimedia
cable	>1M	>2-300k	bajo	Multimedia
FTTH	>20M	>8M	medio	Multimedia
Ethernet	<100M	<100M	bajo	Multimedia
FR	<2M	<2M	medio	Multimedia
ATM	<155M	<155M	alto	Multimedia
Satélite	<64k	<2M	alto	Multimedia
GSM	<19.2k	<100k	bajo	Datos/Audio
GPRS	<64k	<64k	medio	Datos
UMTS	<384k	<384k	alto	Multimedia
LMDS	<100M	<100M	bajo	Multimedia
WiFi	<2M	<2M	medio	Multimedia
WiMax	<384k	<384k	alto	Multimedia
LTE	<100M	<1G	alto	Multimedia

1803



3.3 Los servicios de Internet

El tráfico que circula por la Red depende de los servicios que en ella se ofrecen, estos servicios utilizan aplicaciones que son las interfaces con la que los usuarios interactúan y hacen uso de dichos servicios. Las redes de acceso, los protocolos que operan sobre ellas y los usuarios que las utilizan actúan como modificadores de este tráfico. Como cada uno de los servicios ofrecidos en Internet tiene unas características de tráfico particular, usualmente los estudios para el modelado de tráfico se realizan sobre un tipo de servicio. A continuación se presentarán los servicios principales de Internet (8):

Servicio Web: Se entiende principalmente como la navegación y búsqueda de información, esto incluye cualquier tipo de contenido con las limitaciones de interpretación del cliente (navegador) en el caso de contenidos multimedia.

Servicio de VoIP: Utilizado para la comunicación de voz por medio de las redes de datos, de forma alterna telefonía fija y móvil. Muy usado como una solución para la comunicación intraempresa pero cuyo uso se ha extendido debido al desarrollo de las redes, a la comunicación metropolitana.

Servicio de Videoconferencia: Servicio de video en tiempo real desarrollado para comunicar a personas ubicadas en distintos lugares del planeta. También permite el intercambio de datos, imágenes y sonidos, es interactiva por lo que facilita la comunicación bidireccional.

Servicio de Streaming: Se utiliza para permitir el acceso de contenido multimedia en Internet, en tiempo real, de tal manera que el usuario puede consumir el contenido a medida que se descarga. Suele asociarse al servicio web para el intercambio de audio y video o contenido multimedia de las páginas web.

Servicio de Transferencia de Ficheros: Permite transferir archivos tipo texto y tipo binario entre computadoras conectadas en red. Está soportado en la arquitectura cliente/servidor y generalmente aplicado al uso de recursos compartidos, pues supera las limitaciones de tamaño que imponen los correos electrónicos. Es muy utilizado dentro del servicio web y el servicio **P2P (Peer-To-Peer)**.



Servicio de correo electrónico: Permite a los usuarios compartir mensajes rápidamente, también se utiliza dentro del servicio web, aunque puede funcionar de forma independiente. Por medio de este servicio se puede compartir todo tipo de archivos.

Servicio P2P: Este servicio crea una red virtual entre dos computadoras, permitiendo de esta forma el intercambio de cualquier tipo de archivos entre ellas. Funciona como una conexión punto a punto entre dos computadoras, y tiene características similares a algunos de los servicios mencionados.



4. EJE DE USUARIOS

Ciertamente los servicios, las tecnologías, los protocolos y la arquitectura de la red influyen en el tráfico, sin embargo la variable fundamental son las personas que interactúan a través de Internet haciendo uso de esos servicios, tecnologías, protocolos y arquitecturas. Las personas se pueden clasificar en dos grupos importantes:

Suscriptores: las personas para acceder a Internet hacen uso de una suscripción, en otras palabras contratan con una ISP un ancho de banda o velocidad de acceso, **tabla 10**. Según la velocidad de acceso que contraten con el operador de Internet, los suscriptores pueden ser residenciales, Pymes y grandes empresas, ver **tabla 8**. En este sentido, y como se observa en los distintos reportes que publican los ministerios de las telecomunicaciones en los distintos países, las velocidades más bajas son las que tienen mayor número de suscriptores. Por ejemplo, el boletín trimestral de conectividad presentado por el Ministerio de Tecnologías de la Información y las Comunicaciones, lo clasifica de la siguiente manera (37):

Tabla 8. Participación de suscriptores por segmentos en Colombia²²

SEGMENTO	PARTICIPACIÓN %		
	2 TRIM.-2010	1 TRIM.-2011	2 TRIM.-2011
RESIDENCIAL	86,44%	87,46%	87,15%
CORPORATIVO	12,80%	12,13%	12,46%
INSTITUCIÓN EDUCATIVA	0,47%	0,25%	0,30%
TELECENTRO	0,08%	0,07%	0,06%
ALCALDÍA	0,14%	0,05%	0,01%
INSTITUCIÓN DE SALUD	0,03%	0,01%	0,01%
CAFÉ INTERNET	0,02%	0,01%	0,02%
GUARNICIÓN MILITAR	0,01%	0,01%	0,00%
CENTROS PROVINCIALES DE GESTIÓN AGROEMPRESARIAL	0,00%	0,00%	0,00%
TOTAL	100,00%	100,00%	100,00%

²² (37)

En Latinoamérica los ministerios usan dos tipos de clasificación para agrupar a los suscriptores, una por velocidades de acceso, ver **tabla 10** y otra por tecnologías, ver **tabla 9** (38), siendo esta última la más usada:

Tabla 9. Distribución de suscriptores de Internet por tipo de acceso, 2008²³

Medio de acceso	Suscriptores Diciembre 2009	Suscriptores Marzo 2010	Variación
Acceso Conmutado	50.603	35.171	-30,50%
Acceso dedicado fijo			
xDSL	1.427.934	1.493.039	4,56%
Cable	717.761	734.860	2,38%
WiMAX e inalámbricos	33.919	49.494	45,92%
Otros	35.934	32.295	-10,13%
SUBTOTAL DEDICADO FIJO	2.215.548	2.309.688	4,25%
TOTAL ACCESOS FIJOS	2.266.151	2.344.859	3,47%
Acceso a través de redes móviles(*)	915.280	964.671	5,40%
TOTAL (FIJO+MÓVIL)	3.181.431	3.309.530	4,03%

Tabla 10. Cantidad de suscriptores según rango de velocidad²⁴

Rango de velocidad (Kbps)	Suscriptores a Diciembre 2009	Suscriptores a Marzo 2010	Variación
Menos de 256	97.671	88.931	-8,95%
Entre 256-512	96.344	78.983	-18,02%
Entre 512-1024	678.054	728.179	7,39%
Entre 1024-2048	843.584	908.548	7,70%
Entre 2048-4096	469.569	471.132	0,33%
Mayor a 4096	30.326	33.915	11,83%
Totales	2.215.548	2.309.688	4,25%

²³ (38)

²⁴ (38)

Usuarios: Los usuarios a diferencia de los suscriptores son los que realmente generan el tráfico, son los usuarios los que hacen uso de Internet a través una suscripción, por ejemplo, en un hogar con una suscripción varias personas pueden estar conectadas a la Red, ellos son los usuarios.

Existen dos maneras de estimar los usuarios, una del lado del cliente y otra del lado del servidor. La primera se refiere a los usuarios que hacen uso de una misma suscripción a Internet, esta cantidad es publicada por las entidades oficiales, estatales e internacionales, de Telecomunicaciones a través de sus informes trimestrales o anuales. Estas entidades estiman la cantidad de usuarios con acceso a Internet, a través de encuestas en diferentes segmentos (hogares, empresas y centros colectivos), como resultado se obtiene un factor de multiplicación que representa el promedio de usuarios por suscripción, y un parámetro de ajuste conocido como factor de multiplicidad que permite minimizar la probabilidad de conteo duplicado, porque un mismo usuario tiene la posibilidad de conectarse desde su hogar y su trabajo. Esta estimación tiene varios usos, el más común es para conocer cuántas personas de un país, en promedio, están teniendo acceso a Internet y por tanto cual es la penetración de Internet en ese país (38).

La segunda es hacer la estimación del lado del servidor, esta es un poco más complicada, por lo general se hace por medio de la identificación de logs²⁵. Con esta información se puede saber cuántas veces fue visitada una página. Para identificar cuantas visitas fueron hechas por el mismo usuario, se utilizan cookies que identifican computador, cliente (navegador) y usuario, con lo que permite hacer la estimación de cuantos usuarios son únicos, o en otras palabras cuántas visitas pertenecen al mismo usuario, pues una mismo usuario puede desde un mismo computador usar distintos navegadores, lo que daría como resultado distintos registros de logs. Esta técnica no permite identificar a la persona que usa el computador, por ejemplo, en un café internet,

²⁵ Registro de eventos que sirve para almacenar los datos de qué, quién, cuándo, dónde una aplicación o dispositivo hace una solicitud.

un mismo computador y navegador es usado por muchas personas, que para interpretación del servidor, sería el mismo usuario.

Muchas páginas, utilizan esta técnica para realizar el ranking de páginas más visitadas. Entre ellas se encuentra www.alex.com, www.compete.com y www.quancast.com, existen más pero estas son las más reconocidas y confiables. Las dos últimas publican el número de visitantes únicos que entran a una página, donde ese valor representa *“el número de individuos únicos que han visitado la página en un mes”*²⁶.



²⁶ Tomado de: www.compete.com



5. MODELO CASUAL

Este es un modelo reciente, los primeros resultados obtenidos con este modelo se presentaron en 2006 durante la conferencia sobre electrónica, hardware y comunicaciones ópticas organizada por la **WSAES** (*World Scientific and Engineering Academy and Society*) (39), luego en 2008 se presentaron más avances con el artículo "Approximation to a behavioral model for estimating traffic aggregation scenarios" (40) y a finales de 2009 dentro del grupo de investigación de telemática de la Universidad de Cantabria, España fue presentado como tesis doctoral por Alberto Eloy García G. (8).

Conceptualmente este modelo adopta dentro de sí otros modelos de Teletráfico. Utiliza como su base teórica los modelos *On-Off*, utilizados ampliamente en las redes **ATM** para explicar el comportamiento a ráfagas de las llamadas telefónicas y datos (9), ver **figura 14**, también hace uso de los modelos autosimilares que describen la correlación del tráfico de Internet a distintas escalas de tiempo y por último los modelos **COM** (5) que explican las bases naturales del tráfico debido al uso del protocolo **TCP/IP**.

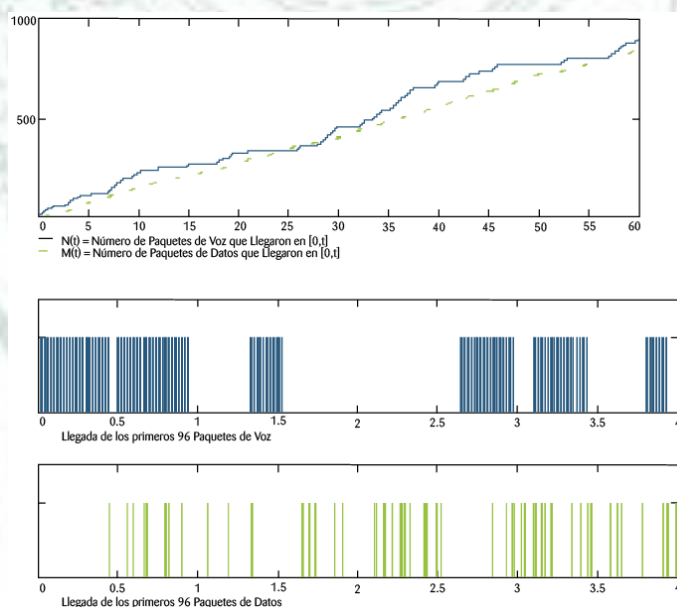


Figura 14. Series de tiempo que representan el tráfico de voz y datos sobre una red local²⁷

²⁷ Alzate Monroy, Marco A. "Introducción al tráfico autosimilar en redes de comunicaciones". En Ciencia, Investigación y Desarrollo. Vol. 6. N° 2. pp. 6-17, 2001

Los modelos *On-Off* adquieren importancia en la década de los 80s, cuando son incluidos por la **ITU** (*International Telecommunication Union*) en distintas recomendaciones para la modelización de tráfico (22). Aparecen como una necesidad de optimizar el uso de las líneas telefónicas en los tiempos de espera de conexión. Se caracterizan por dos estados, uno *On* o estado activo donde se produce la ráfaga y otro *Off* o estado de espera. Este modelo puede aproximarse con una cadena de Markov de dos estados. Sin embargo, los modelos *On-Off* por sí solos no son autosimilares y no pueden describir el comportamiento de dependencia a largo plazo del tráfico de Internet.

Para sobre pasar este inconveniente el modelo **CASUAL** sigue las recomendaciones de la **ITU**, que sugiere para el modelado de tráfico de datos considerar distintos niveles de tráfico (3), en este caso el modelo **CASUAL** analiza los niveles de conexión, sesión y ráfaga, y por tanto toma parámetros de los modelos **COM** que son correlacionados o de autosimilitud. Con esto puede describir la tendencia a corto y largo plazo del tráfico generado por los servicios de Internet. El aporte del modelo **CASUAL** a los modelos de tráfico *On-Off* es que permite asociar el tráfico de Internet a la demanda por parte de los usuarios de los distintos servicios ofrecidos en la Red, **VoIP** (*Voice over Internet Protocol*), Web, Streaming, **FTP**, **P2P**, Juegos, Video y Audio, facilitando de esta forma, el análisis de la demanda de tráfico combinando tres ejes, el eje de usuarios, el eje de servicios y el eje de acceso.

5.2 Ventajas y desventajas

- El modelo **CASUAL** puede describir las tendencias de corto y largo plazo del tráfico de Internet, tanto para una fuente como para múltiples fuentes
- La parsimonia heredada de los modelos *On-Off* permite una interpretación sencilla respecto a otros modelos que describen procesos multifuentes
- Cuantos más niveles de tráfico se agreguen al modelo, mejor estimará el proceso que se pretende modelar, sin embargo aumenta el nivel de complejidad del modelo y se pierde la parsimonia



- Cuando las fuentes tienden a infinito el modelo se hace extremadamente complejo, esto debido a que la técnica de modelado es muy rígida, pues sólo hay dos estados, uno activo donde se transmite y otro de espera donde no se transmite nada, este fenómeno no siempre representa un canal de comunicaciones, que en ocasiones, dependiendo de las aplicaciones o servicios, necesita ser representado con más estados
- Permite estimar en alto grado la aleatoriedad de distintos tipos de tráfico con características independientes, uniformemente distribuidos, de cola pesada y autosimilares
- Combina las tecnologías de acceso, los tipos de servicios y los usuarios ofreciendo de esta forma un escenario más cercano a la agregación del tráfico en los distintos nodos de intercambio de la Red.

5.3. Características del modelo CASUAL

Este modelo considera que el tráfico es el resultado de la interacción de tres ejes, el eje de usuarios, el eje de servicios y el eje de acceso, ver **figura 15**, de ahí su nombre, **CASUAL** (*Cubo²⁸ de Acceso, Servicios y Usuarios de Asignación Libre*). El conjunto de los tres ejes forman el espacio **CASUAL**, donde cada uno de estos ejes modifica el comportamiento del tráfico, y el sentido de las flecha indica hacia donde se produce la agregación de tráfico en los distintos planos de observación.

²⁸ La palabra **Cubo** hace referencia a la naturaleza autosimilar (fractal) del tráfico de Internet, en el espacio **CASUAL**, las unidades en cada uno de los ejes son el tipo de usuarios (residencial, pyme, gran empresa), el tipo de acceso (**ADSL**, Dial-Up, Satélite...) y el tipo de servicios (**VoIP**, **P2P**, Web...), y a su vez, estas unidades son distintos rangos, rangos de usuarios, rangos de velocidades de acceso y rangos de servicios (debido a los protocolos), en este sentido el espacio **CASUAL** está constituido por cubos, en cualquier nivel de agregación y no por puntos como un espacio geométrico normal



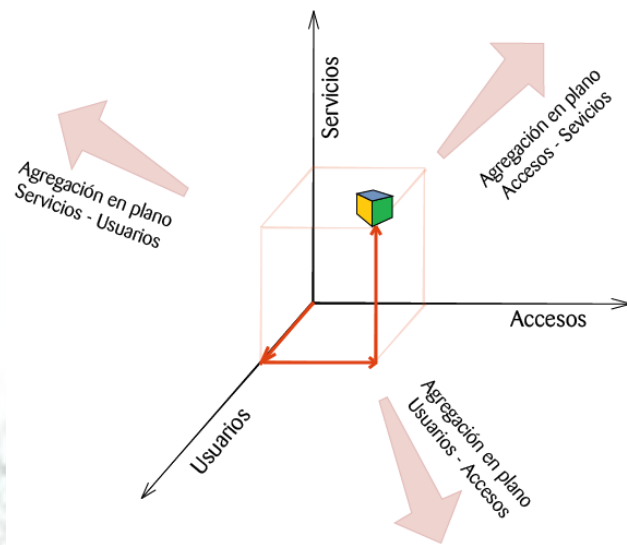


Figura 15. Espacio CASUAL y sus planos de agregación

El eje de usuarios expresa las necesidades de los clientes, sus gustos y sus costumbres; de una manera general los usuarios se pueden agrupar por niveles económicos y laborales como usuarios residenciales, grandes empresas, medianas y pequeñas.

El eje de servicios es el que sirve de interfaz entre el proveedor y el cliente, como VoIP, Streaming, Web, P2P, TVoD, FTP, juego, Video conferencia y otros; estos pueden ser usados de distintas formas y finalidades, depende de las características ambientales, morfológicas y circunstanciales del usuario, lo que influye también sobre el tráfico. El eje de acceso representa el tipo de canal usado para acceder a los servicios, en general se representa con la velocidad máxima que permite el canal, que depende de la tecnología como Dial-Up, ADSL, FTTx²⁹ (*Fiber To The Home*), etc.

Según estas consideraciones, los usuarios, los servicios y los accesos representan las unidades que dividen cada eje, y por esto la combinación de las unidades en el espacio CASUAL se interpretan como un cubo. Ver figura 16.

Este modelo recoge características del tráfico de multifuentes, en particular la multiplexación de múltiples fuentes, esto permite el análisis del tráfico desde la capa de aplicación hasta la capa física y sus distintos niveles de agregación, pero además la dependencia de largo plazo. Esta solución de los modelos *On-Off* considera que los

²⁹ Aunque el acrónimo correcto sería FTTH, se ha colocado intencionalmente FTTx para referir los distintos accesos ofrecidos con fibra óptica, FTTN, FTTC, FTTB, FTTP, FTTD

flujos de paquetes generados por los servicios ofrecidos en Internet, recorren la red pasando por distintos nodos de interconexión, agregándose en cada nodo hasta la fuente, y en general que la dependencia a largo plazo puede estar asociada a los protocolos que implementan para su funcionamiento los distintos servicios.

Se presentan las características principales de los modelos *On-Off*, las características de agregación, distintas distribuciones de probabilidad, el teorema central del límite, algunas características de los modelos multifuentes, de los modelos multinivel y el análisis de las probabilidades en los niveles de sesión y ráfaga.

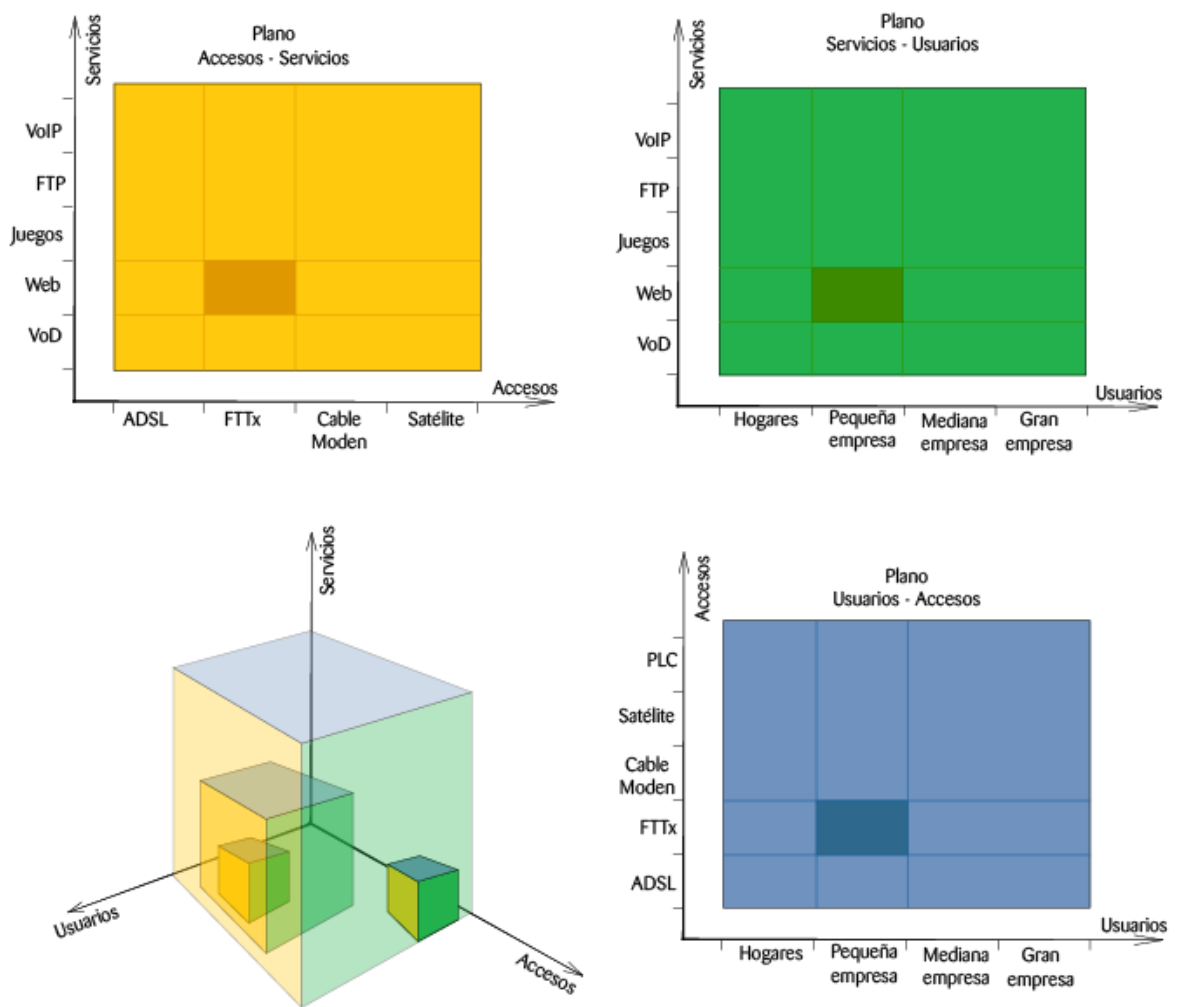


Figura 16. Interpretación del espacio CASUAL

1 8 0 3



5.3.1. Características del modelo ON-OFF en el modelo CASUAL

Para abordar como los modelos *On-Off* pueden usarse para la estimación del tráfico de Internet, es importante entender las distribuciones de cola pesada o decaimiento lento o hiperbólico³⁰, la más conocida es la distribución de Pareto.

En el estudio presentado con el artículo “*Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level*” (41), los investigadores demuestran que la suma de muchos procesos *On-Off* cuyos periodos *On* y *Off* son independientes, pueden representarse con una distribución con decaimiento hiperbólico de tal manera que las series temporales allí representadas son asintóticamente autosimilares, con parámetro de Hurst (H). Tal como se ilustra expresión 5.3.1.1.

$$H = \frac{(3 - \alpha)}{2}, \quad \text{con } 1 > H > \frac{1}{2} \quad 5.3.1.1.$$

En este caso el periodo *On* u *Off* con mayor decaimiento hiperbólico, con α más pequeña, será la cola pesada predominante para el cálculo de H (42).

Este hecho permite analizar la correlación a distintas escalas de tiempo y por tanto la dependencia a largo plazo. Estos modelos LRD presentan una autocorrelación que decae con $k^{-\alpha}$ y que sirve para estimar el grado de dependencia de los sucesos entre el instante de tiempo t y $t+k$, el exponente α de la LRD puede tomar valores en el intervalo $[0,2]$ (42).

Distintas publicaciones a nivel mundial sobre estudios de tráfico hechos en redes LAN sobre tamaño de ficheros, tiempos de transmisión, retardos, etc., demuestran que la autocorrelación decae hiperbólicamente.

Con respecto al tráfico del servicio web, considerando por ejemplo un nivel de sesión con momentos de transmisión *On* y silencios *Off* con una tasa regular y además, si la duración de los periodos *On* presentaran un decaimiento hiperbólico, entonces lo más probable es que el tráfico tenga características autosimilares. Por este motivo, el modelo más usado para explicar este fenómeno de dependencia a largo plazo utiliza la superposición de fuentes *On-Off*.

³⁰ La explicación de este tipo de distribución se hará en una sección posterior



Cuando se usa una distribución de Poisson con decaimiento hiperbólico, esta combinación se conoce como Poisson *On-Off*, y permite no sólo ajustar la media, sino también la varianza. Además debido al comportamiento impulsivo del tráfico de datos, este modelo permite también analizar la dependencia a corto plazo. Este proceso de Poisson *On-Off*, alterna entre un estado *On* y uno *Off*, según una distribución exponencial de medias μ_{on} y μ_{off} respectivamente. La probabilidad P de que el proceso se encuentre activo *On* o en espera *Off* se puede expresar mediante las relaciones 5.3.1.2 y 5.3.1.3, respectivamente (43):

$$P_{on} = \frac{\mu_{on}}{\mu_{on} + \mu_{off}} \quad 5.3.1.2.$$

$$P_{off} = \frac{\mu_{off}}{\mu_{on} + \mu_{off}} \quad 5.3.1.3.$$

El modelo **CASUAL** hace uso de este cálculo para estimar las probabilidades *On-Off* de sesión P_s y de ráfaga P_r . Una observación importante sobre el cálculo de estas probabilidades es, que al usar el valor medio de los periodos *On* y *Off*, la probabilidad de ocurrencia de cada uno de los estados *On* y *Off* se distribuye de manera uniforme en cada uno de los periodos.

Otra de las ventajas que presenta el modelo de Poisson *On-Off* es que permite analizar la variable aleatoria de entrada como una variable aleatoria de Bernoulli y a partir de ahí derivar un cálculo sencillo para la media y la varianza. Sin embargo el concepto más poderoso queda consignado en la ecuación 5.3.1.4, que como se verá más adelante, es la ecuación que presenta el modelo **CASUAL** para calcular el ancho de banda. A continuación se presenta una breve descripción del proceso de Poisson *On-Off*:

“X es una variable aleatoria Bernoulli que toma el valor 1 si el proceso se encuentra en estado On y valor 0 si se encuentra en estado Off. X_n es un proceso que representa el estado del proceso en un instante n.



A_n es un proceso que representa la tasa de transmisión en el instante n . λ representa la velocidad media en el proceso poissoniano. variables que se relacionan mediante la expresión 5.1.3.4.

$$A_n = \lambda X_n \quad 5.3.1.4.$$

A_n son independientes, distribuidos idénticamente, pero su distribución se permite que sea general. Ver ecuación 5.1.3.5 y 5.1.3.6.

$$E[A_n] = \lambda P_{on} \quad 5.3.1.5.$$

$$\text{Var}[A_n] = \lambda^2 (P_{on})(1 - P_{on}) \quad 5.3.1.6.$$

El modelo de Poisson On-Off es un caso particular del proceso de Poisson modulado por Markov, donde se mezcla la sencillez de la conmutación markoviana con la generación poissoniana y λ cambia según el estado en que se encuentre el proceso de Markov' (43).

Es importante tener presente que al ser muchos experimentos de Bernoulli la media y la varianza de 5.3.1.5 y 5.3.1.6 respectivamente, corresponden a una distribución binomial. Un experimento de Bernoulli es el caso más simple de una distribución binomial cuya probabilidad de ocurrir 1 ó 0 es la misma. El experimento de Bernoulli más conocido es el lanzamiento de una moneda equilibrada donde la probabilidad de ocurrir cara o sello es igual en cada lanzamiento.

5.3.2. Distribuciones Estadísticas

El modelo CASUAL usa tres niveles *On-Off*, el de conexión, el de sesión y el de ráfaga, ver figura 17. Para los dos primeros niveles puede usarse la distribución binomial general o la binomial negativa, eso depende del tipo de servicio que se quiera analizar, pero para el nivel de ráfaga necesariamente se usa la distribución binomial negativa.



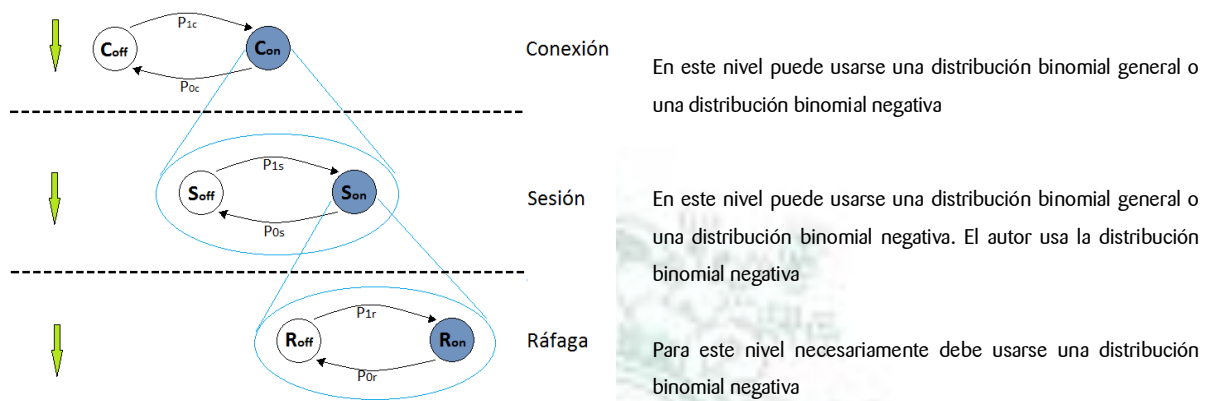


Figura 17. Niveles *On-Off* del modelo CASUAL y sus distribuciones probabilísticas

Se usa binomial negativa para los niveles de sesión y de ráfaga porque para cualquiera de los otros dos casos, binomial normal o binomial positiva la capacidad hallada estaría subestimada (8).

5.3.2.1. Distribución Binomial

Es una distribución de probabilidad de tipo discreta. Mide los éxitos ocurridos en n ensayos de Bernoulli independientes, con probabilidad de éxito fija p entre ensayos.

Su función de probabilidad se puede ver en la ecuación 5.3.2.1.1.

$$f(X) = \binom{n}{x} p^x (1-p)^{n-x}$$

5.3.2.1.1.

Su media se puede ver en la ecuación 5.3.2.1.2.

$$E(X) = np$$

5.3.2.1.2.

Y su varianza se puede ver en la ecuación 5.3.2.1.3.

$$Var(X) = np(1-p)$$

5.3.2.1.3.

En la figura 18 se muestran ejemplos de las distintas formas de la distribución binomial.

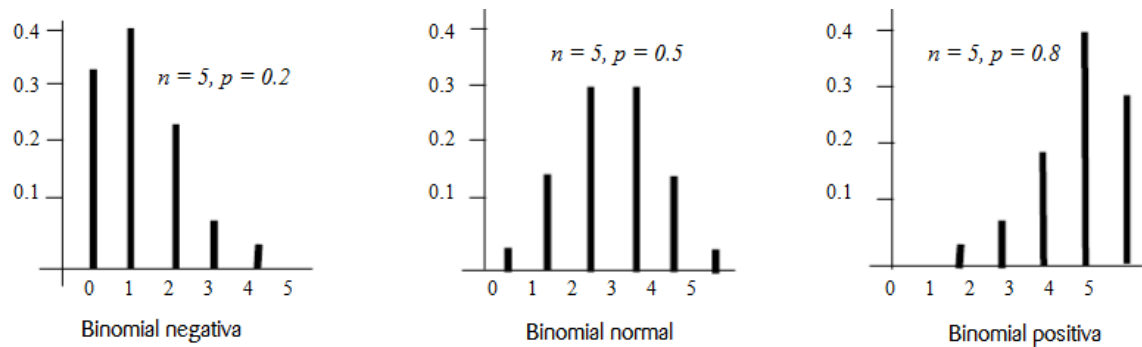


Figura 18. Distintos tipos de distribución binomial³¹

5.3.2.2. Distribución Gamma

Se utiliza generalmente para describir variables aleatorias continuas cuyos sucesos se presentan en mayor densidad al lado izquierdo de la media que al derecho. Se llama así porque la función Gamma es la responsable de la convergencia de la distribución.

Tiene dos parámetros principales: α o parámetro de forma y β o parámetro de alcance o asimetría positiva que desplaza la densidad hacia la izquierda. La más conocida de las distribuciones con este tipo de asimetría es la distribución binomial negativa.

Su función de probabilidad se puede ver en la ecuación 5.3.2.2.1.

$$f(X) = P(X \leq x) = \frac{1}{\beta^{\alpha(\alpha)}} * \int_0^x u^{\alpha-1} * e^{-\frac{u}{\beta}} * du \quad 5.3.2.2.1.$$

Su media se puede ver en la ecuación 5.3.2.2.2.

$$E(X) = \alpha\beta \quad 5.3.2.2.2.$$

Y su varianza se puede ver en la ecuación 5.3.2.2.3.

$$Var(X) = E(X^2) - (E(X))^2 \quad 5.3.2.2.3.$$

³¹ Tomado de: <http://dc95.4shared.com/doc/auQXDrYw/preview.html>

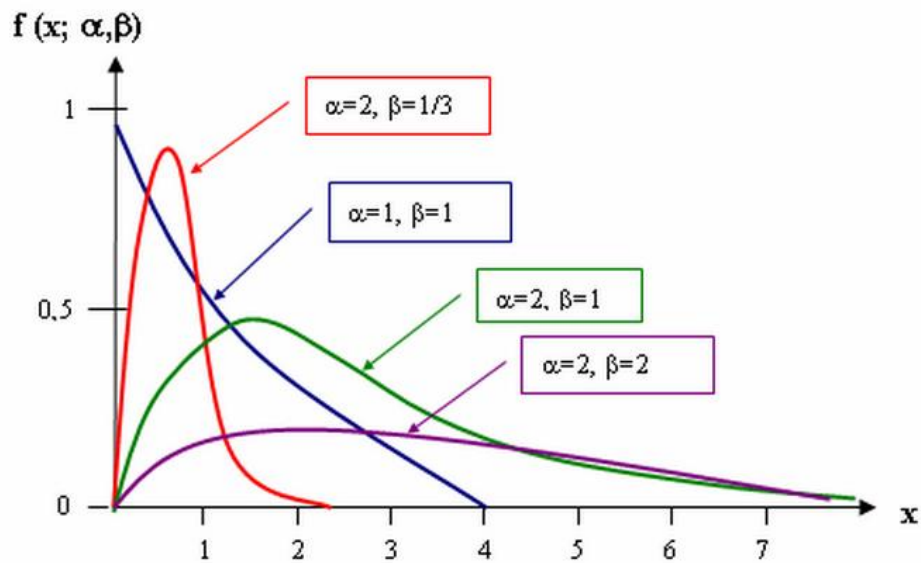


Figura 19. Funções de densidade gamma para distintos pares de parâmetros α y β ³²

5.3.2.3. Distribución Binomial Negativa

Es una distribución de probabilidad que mide la cantidad de experimentos de Bernoulli independientes con probabilidad fija p que se realizan hasta que se obtiene el r -ésimo éxito. Sus parámetros principales son p y r . Esta función es un caso especial de la función Gamma (44).

Su función de probabilidad se puede ver en la ecuación 5.3.2.3.1.

$$f(X) = \frac{(r+k-1)!}{k!(r-1)!} p^r (1-p)^k \quad 5.3.2.3.1.$$

Su media se puede ver en la ecuación 5.3.2.3.2.

$$E(X) = r \frac{1-p}{p} \quad 5.3.2.3.2.$$

Y su varianza se puede ver en la ecuación 5.3.2.3.3.

$$Var(X) = r \frac{1-p}{p^2} \quad 5.3.2.3.3.$$

³² Díaz, Diego. Límite de alerta y de acción para el monitoreo microbiológico ambiental. Tomado de: <http://www.monografias.com/trabajos46/limites-control-ambiental/limites-control-ambiental2.shtml>

5.3.2.4. Distribuciones con decaimiento hiperbólico o de cola pesada en el tráfico web

Muchos estudios se han realizado sobre el tráfico web, el mismo que circula por el puerto 80. Estos estudios han demostrado que el tráfico web tiene una distribución de cola pesada.

En 1998, el tráfico web era el de mayor porcentaje comparado con los otros servicios y aplicaciones que usaban la Red y por esto se consideró que entender las características de este tráfico ayudaría a entender el comportamiento general del tráfico de Internet, la planificación de redes y la evaluación de la calidad de servicio de la misma. Para el estudio “*Heavy-Tailed Probability Distributions in the World Wide Web*” (42), uno de los más importantes de esa época, en el que se analizaron muchas trazas de tráfico web, de una red LAN por un periodo de un mes, tanto del lado del cliente como del lado del servidor. El tráfico generado correspondía a la descarga de distintos tipos de archivos y los datos de tráfico obtenidos eran el nombre del archivo, el tamaño de cada paquete y el tiempo que duraba la transmisión, en el tráfico estaba incluida la sobrecarga que impone el protocolo HTTP.

Uno de los estudios realizados en el artículo fue la prueba de distribución límite, con el objetivo de analizar si los datos tenía varianza infinita, lo que demostraba si era o no una distribución con decaimiento hiperbólico. Las pruebas mostraron que el tráfico cumplía con una de las propiedades de la distribución de colas pesadas, su distribución acumulada complementaria $F(X)$ tenía las siguientes características que se puede ver en la ecuación 5.3.2.4.1.

$$F(X) = 1 - F(X) = P[X > x] \quad 5.3.2.4.1.$$

Que cumple con: (ver la ecuación 5.3.2.4.2.)

$$\frac{d \log F(x)}{d \log(x)} \sim -\alpha \quad 5.3.2.4.2.$$

Donde α es el índice de la cola como se trató en la sección 5.3.1. En general, se refiere a que la distribución posee un comportamiento hiperbólico asintótico, es decir que la cola se comporta de la forma $k^{-\alpha}$. Para que una cola se considere pesada es suficiente con que α esté en el rango $[0,2]$ pero además para que conserve las características LDR



y por tanto de autosimilitud, el rango de α debe reducirse a $[1,2]$, que es el intervalo para los cuales 4.3.1.1 toma valores en el rango $[0.5,1]$, característicos en los datos que tienen autosimilitud .

La distribución con decaimiento hiperbólico³³ o cola pesada más conocida y más usada para los estudios de tráfico de Internet en general, es la distribución de Pareto cuya función de distribución se puede ver en la ecuación 5.3.2.4.3.

$$F(x; \alpha, x_m) = P\{X \leq x\} = \begin{cases} 1 - \left(\frac{x}{x_m}\right)^{-\alpha}, & \text{para } \alpha > 0, x \geq x_m \\ 0 & , \text{para } \alpha > 0, x < x_m \end{cases} \quad 5.3.2.4.3.$$

Donde X es la variable aleatoria, x_m representa el valor mínimo que puede tomar la variable y por tanto es un parámetro de escalado, α es el índice de cola, y $F(x; \alpha, x_m)$ es la función de distribución acumulada.

En la **figura 20**, puede observarse la diferencia entre un decaimiento exponencial y el decaimiento hiperbólico y como a medida que cambia el índice de la cola, la convergencia se hace más lenta, la cola se hace más pesada a medida que α se acerca cada vez más a 1.

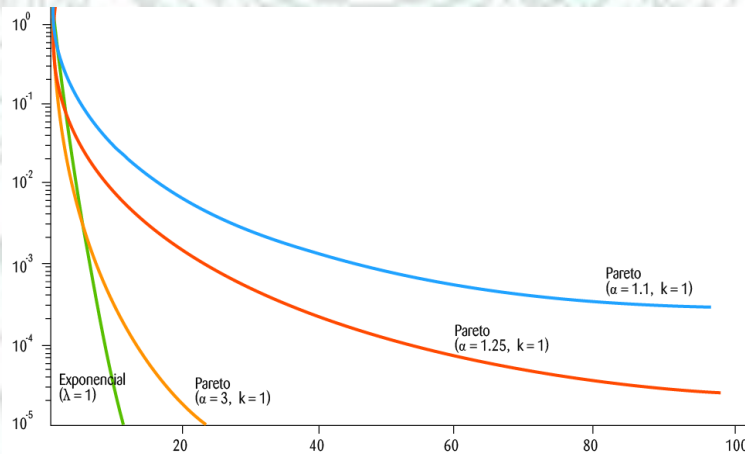


Figura 20. Variación del índice de cola pesada³⁴

³³ Esta suposición se puede hacer porque el decaimiento de la varianza en el tiempo hasta que el sistema alcanza su estado estable, es tan lento (las escalas de tiempo son muy grandes respecto a la muestra), que el valor inicial de la varianza de la muestra es muy grande y esta se puede considerar infinita. En el caso de la mayoría de los servicios de Internet, el tráfico generado nunca llega al estado estable para ninguna escala de tiempo en que se mire. Por este motivo todos aquellos procesos donde ocurre la varianza infinita presentan decaimiento hiperbólico

En el experimento también se estudió el efecto del tráfico agregado, para muestras de distintos tamaños y se observó que se cumplía la teoría de distribuciones estables (42), de tal manera que para los distintos niveles de agregación los experimentos seguían la misma ley potencial y con el mismo índice α , lo que determinó que el tráfico agregado tenía características de varianza infinita³⁵ similares y por tanto decaimiento hiperbólico.

5.3.2.5. Teorema del Límite Central

Este teorema es resultado del trabajo de muchos matemáticos, entre los que se encuentran Laplace, De Moivre, Poisson, Dirichlet, Cauchy, Markov, Liapunov, Lindeberg, Lévy, Feller, Polyá, entre otros. Sin embargo es importante destacar a Laplace quien descubrió este importante teorema, aunque nunca lo demostró rigurosamente y tampoco enunció formalmente, y a Lévy quien demostró por medio de funciones características el trabajo comenzado por Laplace. Su demostración comprobó que el **TLC (Teorema del Límite Central)** podía ser aplicado a distribuciones estables con varianza infinita (44).

Realmente el **TLC** no es un único teorema, sino que está compuesto por distintos resultados sobre el comportamiento del promedio o suma de variables aleatorias. Con él se afirma bajo ciertas hipótesis que la distribución de la suma o promedio de un número grande de variables aleatorias tienden a una distribución normal.

En 1920 Polyá le apodó “Central” haciendo referencia a su importancia dentro de la teoría de las probabilidades y su capacidad para explicar porque tantos fenómenos físicos presentan distribuciones normales o casi normales. Sus resultados pueden ser usados para el contraste de hipótesis, cálculo de intervalos de confianza, estimación de tamaños de muestras, entre otros.

³⁴ STALLINGS, William. Redes de Internet de alta velocidad, rendimiento y calidad de servicio. 2ª ed. Madrid: PEARSON EDUCACIÓN, S.A., 2004. 752 p. ISBN: 84-205-3921-X.

³⁵ La característica ideal es que la varianza sea infinita, de hecho, para las demostraciones matemáticas de esta propiedad se supone varianza infinita. Recibe el nombre de varianza infinita porque literalmente el valor de la varianza es astronómicamente grande con respecto a la media de la muestra

En general, el TLC comprueba que si una muestra es lo suficientemente grande, no importa cuál sea la distribución de la variable aleatoria que se esté estudiando, la distribución de su media muestral, o en otras palabras, la distribución de las medias de las muestras tomadas sobre la variable aleatoria, será aproximadamente una normal. No sólo eso, sino que la media muestral coincidirá con la media de la variable aleatoria, y la desviación típica de la media muestral será aproximadamente el error estándar (45).

Como resultado de lo anterior, **la media de una variable aleatoria con una distribución cualquiera contiene la información suficiente para estudiar el comportamiento de la variable aleatoria** (44). Este resultado es muy importante y de mucha utilidad para el estudio de distribuciones estables, por ejemplo cuando la distribución de una variable aleatoria cualquiera k , su curva hacia la derecha decae de la forma $k^{-\alpha}$ con $0 < \alpha < 2$, con varianza infinita. Como es el caso del tráfico de Internet y las distribuciones con LDR.

5.4. Parámetros principales del modelo CASUAL

Este modelo se apoya en el modelo *On-Off* multinivel binomial que permite estudiar la agregación de varios flujos de tráfico independientes en la salida de cualquier dispositivo concentrador, por ejemplo a la entrada de una red troncal, donde se puede observar el tráfico como unas secuencias de paquetes en ráfagas, separados por silencios, sin importar la naturaleza de los paquetes que están transitando. Esto se logra por medio de varios niveles de estados *On-Off*, cada estado con una probabilidad de ocurrencia.

La aplicación de este modelo implica determinar el número de niveles necesarios donde se ubicaran cada par de estados *On-Off* y las probabilidades de los estados *On* y *Off* respectivamente. Existen análisis donde se usan 5, 6 y hasta 7 niveles de profundidad, obteniendo precisiones más altas, pero a su vez se pierde la parsimonia buscada al modelar tráfico, pues se incrementa la dificultad matemática y el costo computacional. El modelo CASUAL opta por usar tres niveles. Como se observa en la **figura 21**.

El nivel de conexión permite modelar el comportamiento de accesos consecutivos a Internet. En el caso de accesos con banda ancha, la conexión se establece cuando el computador accede a la red.



El nivel de sesión permite modelar el tiempo entre dos sesiones consecutivas, entendiendo sesión como la descarga de una única página web.

El nivel de ráfaga permite modelar los patrones de tráfico dentro de una sesión como el tiempo de llegada de la ráfaga de paquetes pertenecientes al servicio o tiempo de llegada de objetos. Ver **figura 22**.

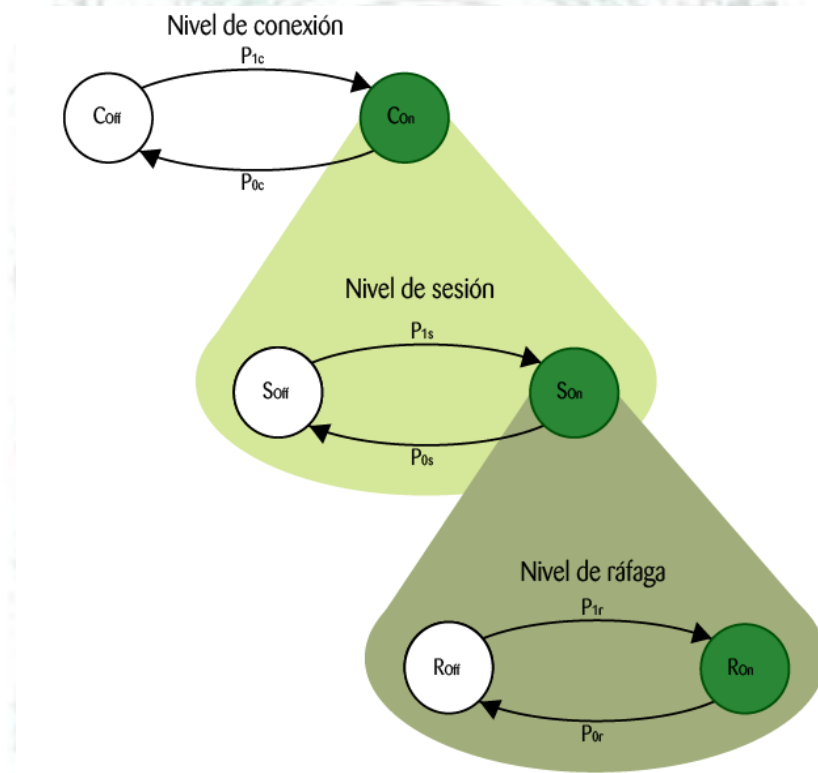


Figura 21. Niveles del modelo CASUAL y los pares de estados *On-Off*

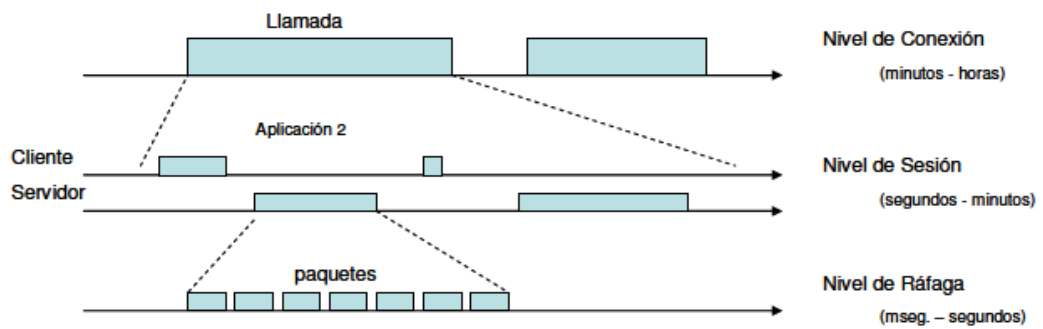


Figura 22. Niveles On-Off del modelo CASUAL en el eje de servicios³⁶

Es importante recordar que el modelo **CASUAL** está compuesto por tres ejes, el eje de usuarios, el eje de servicios y el eje de acceso. Estos tres ejes son inter dependientes, debido a que el tráfico es producto del tipo de servicio, los protocolos que le permiten funcionar y de las aplicaciones que hacen uso de ellos. También es fruto de las distintas formas de uso, finalidades, cultura, situaciones circunstanciales, ambientales y posibilidades de acceso de los usuarios al momento de utilizar un servicio en concreto. Todas estas relaciones modifica el tráfico que circula por la Red.

A continuación se describirá como el eje de servicios afecta el tráfico, específicamente en la descarga de páginas por medio del servicio Web:

- Está soportado por la arquitectura Cliente/Servidor, lo que genera un conjunto de paquetes de control para la sincronización y administración de los paquetes enviados y recibidos
- Como ya se ha mencionado, el servicio web es orientado a conexión, lo que significa que entre el cliente y el servidor se establece un circuito virtual que comienza con una orden sincronización **SYN** y termina con una orden de finalización **FIN**

³⁶ García Gutiérrez, Alberto Eloy. "Contribución al desarrollo de herramientas estratégicas para el diseño, dimensionado y evaluación de redes de telecomunicación de banda ancha". Tesis doctoral. España : Universidad de Cantabria, 2009

- El tráfico está regido por el conjunto de protocolos TCP/IP y Ethernet, que convierten la información en paquetes, por tanto los datos viajan a través de la red en bloques pequeños
- Los paquetes deben viajar por distintos nodos de la Red, a través de distintos dispositivos, que los encolan, y redireccionan hacia su destino. Muchos de estos dispositivos concentran el tráfico, lo que se conoce como agregación, de tal manera que unos cuantos **kbs** (kilobits) de tráfico generados por un hogar pueden llegar a ser, después de ser agregados en una red troncal, unos cientos de **Gbs** (Gigabits)
- El protocolo de interés es **HTTP** que rige la transferencia de páginas web. La versión 1.2 de **HTTP**, permite a los navegadores web abrir la conexión cuando envían por primera vez una solicitud **HTTP** y dejar abierta la conexión para todas las demás solicitudes derivadas en la misma conexión, hechas al mismo servidor
- Los paquetes, no viajan eternamente por la red, tienen un tiempo de vida suficiente para llegar a su destino, de lo contrario son desechados

De las características mencionadas se pueden extraer por un lado, que debido al protocolo **TCP** existe un nivel de conexión, por el otro lado, debido al protocolo **HTTP**, existe un nivel de ráfaga. Se considera una sesión el tiempo que el usuario demora en descargar una página y visualizarla.

Los parámetros para cada uno de los niveles que componen el modelo:

“Nivel de conexión:

- *Número de conexiones por hora (α): Número de conexiones por hora y en un día normal que realiza un usuario del tipo contemplado*
- *Duración de la conexión (\bar{T}_s): Tiempo que el usuario permanece conectado al ISP de manera continuada*
- *Velocidad del acceso (V_{ac}): Ancho de banda de la línea de acceso a Internet*

Nivel de sesión:



- *Tiempo de transferencia ($T_{página}$): Tiempo que tarda una página Web solicitada en ser descargada*
- *Tiempo de visionado ($T_{thinking}$): Es el tiempo durante el cual el usuario no realiza otra petición de página*
- *Tiempo entre sesiones: Tiempo que pasa entre dos sesiones consecutivas pedidas por el usuario. Normalmente suele ser igual a $T_{página} + T_{thinking}$*

Nivel de ráfaga:

- *Tiempo entre ráfagas (T_{idle}): Tiempo que pasa entre la descarga de dos objetos consecutivos pertenecientes a una página Web*
- *Duración de ráfaga (T_{on}): Tiempo de descarga del objeto*

El modelado del nivel de sesión queda definido prácticamente, por el número de peticiones de páginas que realiza el usuario. Dicho número es necesario para el cálculo del tiempo de duración de la sesión, aunque no de forma determinante ya que va a variar enormemente en función de: la velocidad del acceso a Internet, del tamaño de las páginas solicitadas, e incluso del estado de la red.

Por otro lado, en el nivel de ráfaga se realiza la descarga de cada uno de los objetos in-line pertenecientes a cada uno de los “marcos de página” principales, y que representan en definitiva la página Web.

Ambos niveles son semejantes y puede modelarse su agregación mediante binomiales negativas, ya que los dos niveles presentan el comportamiento de ráfagas de naturaleza muy similar, lo cual implica la consabida característica de autosimilaridad de este tipo de tráfico.

Para su cálculo, se va a partir del número de usuarios de un determinado tipo, que hacen un uso similar del servicio Web, y que, con muy alta probabilidad, en media presentan valores muy similares en el T_{idle} (tiempo entre ráfagas) y el T_{on} (duración de la ráfaga).



Es importante notar que no se necesita conocer el tiempo exacto de duración de las ráfagas y de los tiempos de espera, sino la relación entre ellos, por lo que se puede utilizar sus sumas dentro de una sesión para calcular las probabilidades correspondientes (utilizando $T_{página}$ y $T_{thinking}$).

Aquí solo se va a mostrar la deducción de los valores correspondientes a los diferentes parámetros, comenzando por las probabilidades de activación de cada uno de los tres niveles. Al nivel de sesión se calculan como se puede ver en las ecuaciones 5.4.1. y 5.4.2. respectivamente:

$$P_{1s} = \frac{T_{página}}{T_{página} + T_{thinking}} \quad 5.4.1.$$

$$P_{0s} = \frac{T_{thinking}}{T_{página} + T_{thinking}} \quad 5.4.2.$$

Para el nivel de ráfaga: (ver las ecuaciones 5.4.3. y 5.4.4.)

$$P_{1r} = \frac{T_{on}}{T_{on} + T_{idle}} \quad 5.4.3.$$

$$P_{0r} = \frac{T_{idle}}{T_{on} + T_{idle}} \quad 5.4.4.$$

La probabilidad de conexión se obtiene como se puede ver en la ecuación 5.4.5.

$$P_{1c} = \frac{\alpha * T_s}{3600} \quad 5.4.5.$$

A partir de estas probabilidades, se calcula el número de usuarios que están en una conexión multiplicando el valor de la población por el de P_{1c} (ver ecuación 5.4.6.):

$$N_c = pop * P_{1c} \quad 5.4.6.$$

Una vez se tiene ese valor, se calcula el número de usuarios que están en una sesión simultáneamente (ver ecuación 5.4.7.):

$$N_s = \frac{N_c * P_{1s}}{(1 - P_{1s})} + \gamma \sqrt{\frac{N_c * P_{1s}}{(1 - P_{1s})}} \quad 5.4.7.$$

A continuación se calcula el número de usuarios en una ráfaga (ver ecuación 5.4.8.):

$$N_r = \frac{N_s * P_{1r}}{(1 - P_{1r})} + \gamma \sqrt{\frac{N_s * P_{1r}}{(1 - P_{1r})}} \quad 5.4.8.$$

Finalmente, conocida la velocidad de ráfaga (la velocidad del acceso), el ancho de banda agregado correspondiente al servicio Web se obtiene como (ver ecuación 5.4.9.):

$$BW_{Web} = N_r * V_{ac} \quad 5.4.9.$$

El cálculo de las capacidades efectivas, asociadas a una conexión (correspondiente a un servicio utilizado por un determinado tipo de usuario conectado con una tecnología de acceso dada), se resuelve sin la necesidad de conocer el número de transiciones de un estado a otro para cada nivel temporal, es decir, el número de "objetos" (en general: conexiones, sesiones, ráfagas) intercambiados, sino que basta con conocer, estimar o fijar las probabilidades asociadas a cada estado. **Estos valores son determinados a partir de los tiempos totales medios de permanencia en cada estado.**

El concepto de agregación de tráfico, simplifica el modelado del escenario multiservicio que supone el acceso a Internet de un único usuario, dado que las características del tráfico asociado a una sesión depende del tipo de servicio y aplicación" (8).

El valor gamma γ en los en las ecuaciones 5.4.7 para el nivel de sesión y 4.4.8 para el nivel de ráfaga es una constante que se calcula con (ver ecuación 5.4.10.):



$$\gamma = \sqrt{-2 \ln(\varepsilon) - \ln(2\pi)} \quad 37 \quad 5.4.10.$$

En donde γ actúa como un factor corrector de la desviación estándar en el flujo que ocasiona el tráfico agregado.

La agregación de tráfico que se presenta en 5.4.9 es válida para fuentes moduladas con Markov, que consideran la superposición de N fuentes, cuyo valor se puede calcular con (ver ecuación 5.4.11.):

$$C = \sum_{i=1}^N C_i \quad 5.4.11.$$

Donde C es la capacidad equivalente total y C_i es la capacidad de cada fuente agregada.

5.5. Interpretación de la formulación del modelo CASUAL

Aunque el autor del modelo **CASUAL** es claro al explicar la formulación usada, se añadirá una breve ampliación a esta:

El producto a la derecha de la ecuación 5.4.9, es la conclusión del modelo **CASUAL**, su resultado final, donde se intersecan los tres ejes que forman el modelo, el eje de servicios, el eje de usuarios y el eje de acceso. Esta ecuación se compone de N_r (usuarios en ráfaga), valor que resulta de la intersección o combinación de los ejes de usuarios y servicios, representa el número de usuarios que se encuentran simultáneamente generando tráfico, o mejor dicho, son los usuarios que se encuentran recibiendo simultáneamente paquetes de datos. El V_{ac} (velocidad de acceso) corresponde al eje de acceso, y representa la velocidad a la que el usuario está accediendo a Internet, y por tanto la capacidad de ancho de banda que el suscriptor tiene contratada con la ISP.

Debido a que este modelo está regido principalmente por las características de la distribución binomial, se puede pensar en la ecuación 5.4.9, como se presenta en 5.3.1.4., un proceso de Bernoulli, que describe los estados *On* cuando ocurre un éxito y los estados *Off* cuando ocurre un fracaso y donde N_r representa los éxitos o el estado *On* de la ráfaga. BW_{Web} es a la luz de 5.3.1.4 un proceso que representa la tasa de

³⁷ El autor toma esta fórmula de [36] donde ε es un factor de error debido a las pérdidas por sobrecarga en buffer en el servidor. Los valores típicos para γ son 1,2 y 3, por lo general se usa el valor de 2 para evitar la sobre estimación.

transmisión en un instante y V_{ac} la velocidad media de ocurrencia de éxitos, en este caso recepción de paquetes.

Las ecuaciones 5.4.7 y 5.4.8, corresponden a los niveles de sesión y de ráfaga respectivamente, se modelan mediante distribuciones binomiales negativas, y son las encargadas de imponer las características autosimilares al modelo. Estas explican tres características básicas del tráfico a ráfagas de Internet: después que se establece la conexión y se recibe el primer paquete, es de esperar que lleguen más de forma consecutiva, lo mismo ocurre con los silencios; en el proceso de agregación de fuentes se puede obtener el número de éxitos simultáneos dentro de un grupo de procesos estadísticos independientes en niveles temporales consecutivos como son el de sesión y el de ráfaga, y la pendiente causada por la superposición de fuentes, cambia a medida que aumenta el número agregado de fuentes, de tal manera que el crecimiento de la capacidad equivalente debida a la agregación no crece linealmente. Este fenómeno no se puede garantizar si la distribución es binomial positiva (8).

En la fórmula 5.4.6 el nivel de conexión, se modela mediante binomial positiva, ya que considera un estado de conexión permanente, en el cual, una vez establecida la conexión pasa mucho tiempo antes de escuchar el primer silencio. En esta ecuación, pop es el número de usuarios potenciales. En el nivel de conexión serán los usuarios de Internet, en el nivel de sesión serán los usuarios en conexión y en el nivel de ráfaga serán los usuarios en sesión.

La fórmula 5.4.5., se utiliza para calcular la probabilidad de conexión P_{1c} , y conserva la misma forma que la intensidad de tráfico, entendida como el promedio de llamadas realizadas simultáneamente durante un periodo de tiempo, en este caso un Erlang (1 Erlang = 3600 seg), (ver ecuación 5.5.1.):

$$Erlang = \frac{N_{llamadas} * ACHT(seg)}{3600 seg} \quad 5.5.1.$$

Donde: $N_{llamadas}$, es el número de llamadas, $ACHT$ (*Average Call Holding Time*) es el promedio de duración de cada llamada, normalmente dado en segundos y 3600 es el periodo de tiempo analizado. Las equivalencias entre la formulación de 5.4.5. y 5.5.1. es



la siguiente, (ver ecuación 5.5.2):

$$\alpha = N_{llamadas} \text{ y } T_s = ACHT(\text{seg}) \quad 5.5.2.$$

El valor de 3600 en las dos formulaciones equivale al periodo de tiempo analizado, una hora contada en segundos (2).

Sin embargo para hallar la probabilidad de conexión P_{1c} , en este estudio se ha seguido un camino diferente pero conservando el principio conceptual, “el número de usuarios conectados a Internet en un instante de tiempo”. La probabilidad de que una persona esté conectada puede depender de muchos factores circunstanciales, culturales y sociales, lo que dificultaría el cálculo si se estudiara cada usuario de Internet de forma independiente y teniendo en cuenta todas esas variables que afectan su conexión. Por tanto se ha elegido un camino más sencillo pero igualmente válido, calcular la probabilidad de la forma convencional, casos favorables sobre casos posibles, o en nuestro estudio, el número de horas promedio que los usuario de Internet de un país permanecen conectados en el mes sobre el número de horas en promedio que tiene un mes, como se muestra a continuación, (ver ecuación 5.5.3):

$$P_{1c} = \frac{L_{conexiones} \text{ (horas)}}{720 \text{ (horas)}} \quad 5.5.3.$$

Donde $L_{conexiones}$ es el número de horas promedio que un usuario permanece online de forma continuada en un mes, este dato es publicado anualmente por *Comscore* en sus informes sobre el estado de Internet. Es importante tener en cuenta que este valor promedio incluye todos los sesgos, desde los usuarios que nunca se conectan, los usuarios que van a salas de Internet, los que se conectan desde su casa y los que se conectan desde el trabajo y su validez se apoya en el TLC. De ningún modo este valor puede ser comparado con un solo individuo, pues los hábitos de consumo de este individuo no representan los hábitos de consumo de toda la población.

Las expresiones 5.4.1, 5.4.2, 5.4.3. y 5.4.4, son posibles porque se considera que los estados *On-Off* son IID (*Independientes e Idénticamente Distribuidos*) y se alternan, característica heredada de los procesos de Poisson, que permiten usar la media de los estados *On* u *Off* para conocer su probabilidad de ocurrencia, como se analizó en 5.3.1.2. y 5.3.1.3. Sin embargo estas probabilidades dependen de los valores de $T_{página}$, $T_{thinking}$, T_{on} , y T_{idle} , que se obtienen de manera experimental con las trazas de tráfico de entrega y llegada de paquetes tanto en el lado del servidor como del lado del cliente. Estos valores debieron ser reinterpretados obedeciendo a la naturaleza indirecta de los datos obtenidos de forma gratuita a través de Internet y que soslayan la imposibilidad de un ambiente estratégico de hacer mediciones directas sobre equipos.

Nivel de sesión:

$T_{página}$, tiempo que tarda una página Web solicitada en ser descargada se calculó usando una herramienta de análisis suministrada gratuitamente desde el sitio web www.pingdom.com, que realiza el proceso de descarga de la página solicitada y entrega como resultado, tamaño de descarga, tiempo de descarga, número de objetos descargados, latencia, entre otros.

$T_{thinking}$, es el tiempo durante el cual el usuario no realiza otra petición de página, o se encuentra visionando el contenido de la página solicitada. Se calcula con el tiempo medio en segundos que permanece un usuario visitando la página, dividido por el número de páginas que solicita dentro del sitio menos el tiempo que demora la descarga, (*ver ecuación 5.5.4.*):

$$T_{thinking} = \frac{T_{onsite} \text{ (seg)}}{P_{page/user}} - T_{página} \quad 5.5.4.$$



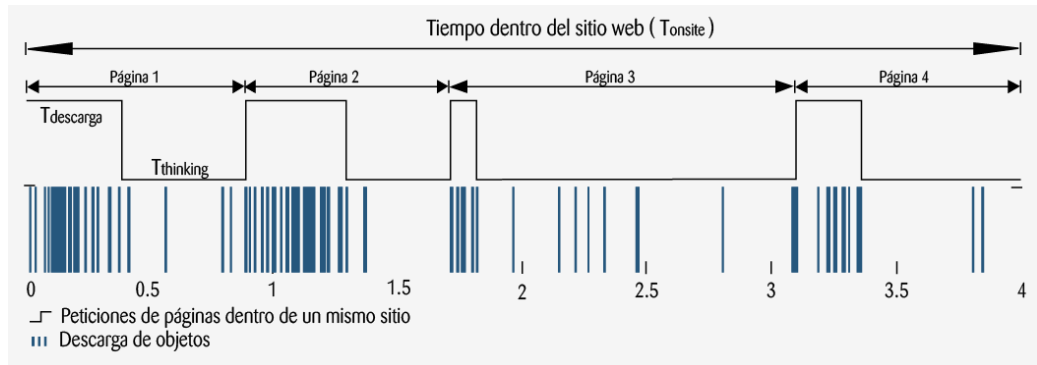


Figura 23. Interpretación de los parámetros en el nivel de sesión

Donde T_{onsite} es el tiempo en segundos que un usuario permanece dentro de un sitio web y n es el número de páginas que consulta dentro del sitio. Esta información también se obtiene de forma gratuita por medio del sitio web www.alexa.com.

Nivel de ráfaga:

T_{idle} , tiempo que pasa entre la descarga de dos objetos consecutivos pertenecientes a una página Web, o latencia del objeto. En este caso se entiende el proceso de descarga de los objetos de forma secuencial, de tal manera que el tiempo de llegada al cliente que separa a dos objetos ubicados en un mismo servidor, es el tiempo que se demoran viajando a través de la misma red.

T_{on} , tiempo de descarga del objeto o el tiempo de recepción del objeto en el cliente. Este parámetro se halla teniendo en cuenta el tiempo total de descarga del objeto que incluye, la parte del objeto que se encuentra en la red y la parte del objeto que se está recibiendo en el cliente, el tiempo utilizado por el servidor para entregar el objeto se considera pequeño respecto a los otros dos y se puede despreciar, menos el número de paquetes que aún se encuentran en la red, (ver ecuación 5.5.5. y 5.5.6.).

$$T_{on} = \frac{T_{descarga} (seg)}{N_{objetos}} - \frac{T_{descarga} (seg)}{N_{objetos}} * T_{idle} \quad 5.5.5.$$

$$T_{on} = \frac{T_{descarga} (seg)}{N_{objetos}} * (1 - Latencia) \quad 5.5.6.$$

Donde $T_{idle} = Latencia$ como ya se había mencionado y $N_{objetos}$ es el número total de objetos solicitados por el cliente para hacer posible la visualización de la página.

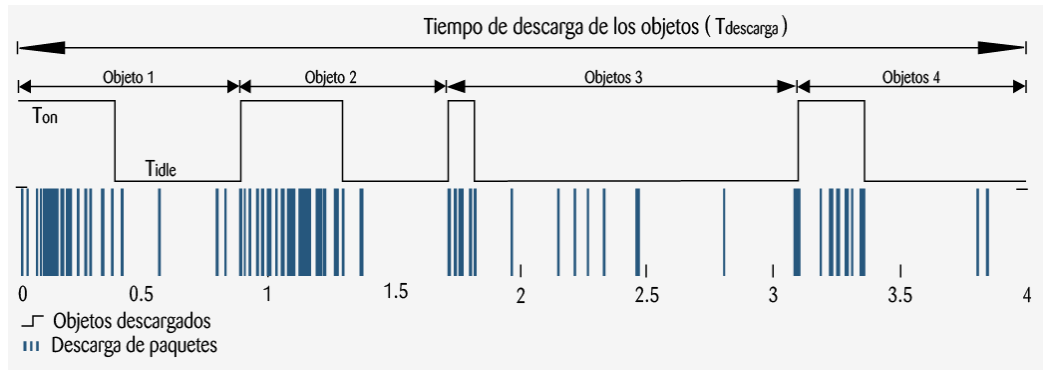


Figura 24. Interpretación del nivel de ráfaga

Se usa la Latencia como un valor límite del número de paquetes del objeto que pueden estar circulando por la red, este número siempre debe ser menor al número de paquetes que se están recibiendo en el cliente, esta característica es inherente al QoS (*Quality of Service*), de hecho, lo que buscan los CDNs (*Content Delivery Network*) es disminuir Latencia cada vez más para que la experiencia que tengan los usuarios sea mejor; el ideal sería *Latencia* = 0. Y por esto se usa la Latencia como un valor limitador de T_{on} .

5.6. Resultados de la implementación del modelo

A continuación se muestran los resultados obtenidos en la implementación del modelo presentado. Primero se muestra el cálculo de los parámetros y la capacidad requerida para una fuente de algunas de las páginas web comunes y más consultadas de los países de la región latinoamericana en estudio. Luego, tomando como referencia una de las páginas, se calculan los parámetros y la capacidad equivalente para N fuentes, en cada uno de los países, también se calcula la capacidad equivalente para N fuentes para distintos años en estos países y por último se usa el cálculo anterior y a partir el porcentaje de tráfico web que representa en cada año se hace una estimación del tráfico total que transita por la red.

5.6.1. Parámetros y capacidad requerida de las páginas más relevantes

En la **tabla 11**, se muestran algunos parámetros del modelo CASUAL y el cálculo de la capacidad requerida al descargar algunas de las páginas más consultadas y comunes en los diferentes países de la región latinoamericana que se encuentran en estudio.

Al principio de la **tabla 11** se encuentran los tres ejes del modelo **CASUAL**, el eje de usuarios, servicio y Accesos, luego al final se encuentran los tres niveles *On-Off*, **Nc** (*Nivel de conexión*), **Ns** (*Nivel de sesión*), y **Nr** (*Nivel de ráfaga*). Por último el cálculo de la capacidad demandada cuando un usuario accede a cada una de las páginas mostradas.

Tabla 11. Capacidad demandada al descargar una página web

Página	Parámetros y capacidad												
	Usuarios	Servicio Web	Acceso Kbps	Tonsite	Page/User	Tdescarga	Nobjetos	latencia	Tonline	Nc	Ns	Nr	Capacidad Kbps
youtube	1	100%	1024	1017,0	12,81	0,680	32	0,12	25,7	0,0357	0,0354	0,1541	157,83
google	1	100%	1024	735,6	11,72	0,399	12	0,12	25,7	0,0357	0,0305	0,1798	184,08
hotmail	1	100%	1024	478,2	6,21	0,240	10	0,12	25,7	0,0357	0,0212	0,1260	129,04
yahoo	1	100%	1024	520,8	6,86	1,060	33	0,12	25,7	0,0357	0,0455	0,2177	222,92
blogspot	1	100%	1024	383,4	4,86	1,900	33	0,12	25,7	0,0357	0,0602	0,3444	352,67
wikipedia	1	100%	1024	266,4	3,88	1,910	16	0,12	25,7	0,0357	0,0649	0,5337	546,52
facebook	1	100%	1024	1488,0	12,9	0,791	11	0,12	25,7	0,0357	0,0316	0,2750	281,65

En la **tabla 12** se hace el cálculo para Colombia, teniendo en cuenta el porcentaje de usuarios y la velocidad de acceso promedio.

Tabla 12. Capacidad requerida por páginas web en Colombia

Página	Parámetros y capacidad							Capacidad Kbps
	Usuarios	% Usuarios	Servicio web	Acceso Kbps	Nc	Ns	Nr	
youtube	25.000.000	0,60%	100%	2176	5.354,17	59,86	15,44	33.589,12
google	25.000.000	0,50%	100%	2176	4.461,81	39,23	15,75	34.275,54
hotmail	25.000.000	1,40%	100%	2176	12.493,06	51,56	15,10	32.855,23
yahoo	25.000.000	0,40%	100%	2176	3.569,44	64,76	23,07	50.193,16
blogspot	25.000.000	1,10%	100%	2176	9.815,97	273,38	136,91	297.921,81
wikipedia	25.000.000	0,90%	100%	2176	8.031,25	260,13	257,90	561.192,31
facebook	25.000.000	0,80%	100%	2176	7.138,89	63,33	44,96	97.825,58

Es interesante hacer varias observaciones, los usuarios usados están del lado del cliente, este dato generalmente es publicado por los ministerios de telecomunicaciones de cada gobierno, la ITU, la CIA y en el sitio www.internetworldstat.com.

Sólo se está considerando el servicio web, más exactamente la operación de entrar y visualizar las páginas mencionadas, los otros servicios de Internet no están incluidos en

el cálculo de la capacidad C mostrada en la **tabla 12**. Para hacerlo, es necesario considerar que el servicio web representa aproximadamente entre el 20% el 17% del tráfico de Internet de 2011, según lo reporta *Akamai*, *Telegeography* y *Cisco* en sus diferentes informes anuales. Con esta información se podría realizar la siguiente operación, (*ver ecuación 5.6.1.1. y 5.6.1.2.*).

$$C_{Facebook} = \frac{98 \text{ (Mbps)}}{0.17} = 576,47 \text{ Mbps} \quad 5.6.1.1.$$

$$C_{Facebook} = \frac{98 \text{ (Mbps)}}{0.20} = 490 \text{ Mbps} \quad 5.6.1.2.$$

Que representaría aproximadamente la capacidad media requerida para atender la demanda que los usuarios de Internet en Colombia hacen de todos los servicios que ofrece Facebook.com en su sitio, videos, chat, email, fotos, juegos, entre otros.

Se puede hacer el ejemplo con el caso Google, aprovechando que tres de las páginas de la **tabla 12** pertenecen a la empresa, **www.youtube.com**, **www.google.com** y **www.blogspot.com**, de tal manera que, aproximadamente, la capacidad media requerida para atender la demanda de los usuarios en Colombia de los servicios que ofrece Google en sus diferentes sitios web es, (*ver ecuación 5.6.1.3. y 5.6.1.4.*).

$$C_{Google} = \frac{(34+34+298) \text{ Mbps}}{0.17} = 2153 \text{ Mbps} = 2.1 \text{ Gbps} \quad 5.6.1.3.$$

$$C_{Google} = \frac{(34+34+298) \text{ Mbps}}{0.2} = 1830 \text{ Mbps} = 1.79 \text{ Gbps} \quad 5.6.1.4.$$

Este cálculo es tentativo y puede ser mayor, pues sólo se han tenido en cuenta 3 de los sitios web que pertenecen a la compañía.

Otro estudio se puede hacer usando los usuarios del lado del servidor para esas páginas reportados por *compete.com* y los porcentajes de usuarios de cada país reportados por *www.alexa.com*, como se observa en la **tabla 13**.



Tabla 13. Capacidad con usuarios del lado del servidor

Página	Parámetros y capacidad							
	Usuarios	% Usuarios	Servicio	Acceso Kbps	Nc	Ns	Nr	Capacidad Kbps
youtube	149.077.396	0,60%	100%	2176	33.045,49	319,28	63,86	138.962,92
google	174.124.771	0,50%	100%	2176	32.164,71	234,47	72,29	157.313,27
hotmail	78.696.438	1,40%	100%	2176	40.703,55	149,82	36,64	79.724,34
yahoo	154.182.513	0,40%	100%	2176	22.784,75	358,56	102,84	223.781,41
blogspot	66.375.041	1,10%	100%	2176	26.974,08	717,29	337,66	734.753,05
wikipedia	88.824.929	0,90%	100%	2176	29.534,29	903,24	846,96	1.842.974,10
facebook	168.599.429	0,80%	100%	2176	49.830,50	381,17	229,36	499.082,69

Con la información consignada en la **tabla 13**, también se puede hacer el cálculo siguiendo la lógica aplicada en $C_{Facebook}$ y C_{Google} , (ver ecuación 5.6.1.5. y 5.6.1.6.).

$$C_{Facebook} = \frac{499 \text{ (Mbps)}}{0.17} = 2935,29 \text{ Mbps} = 2,9 \text{ Gbps} \quad 5.6.1.5.$$

$$C_{Facebook} = \frac{499 \text{ (Mbps)}}{0.2} = 2495 \text{ Mbps} = 2,5 \text{ Gbps} \quad 5.6.1.6.$$

Y para *Google* se tiene, (ver ecuación 5.6.1.7. y 5.6.1.8.).

$$C_{Google} = \frac{(139+157+735) \text{ Mbps}}{0.17} = 6064,7 \text{ Mbps} = 6,06 \text{ Gbps} \quad 5.6.1.7.$$

$$C_{Google} = \frac{(139+157+735) \text{ Mbps}}{0.2} = 5155 \text{ Mbps} = 5.2 \text{ Gbps} \quad 5.6.1.8.$$

5.6.2. Parámetros y capacidad requerida para N fuentes

En la **tabla 14** se muestran los parámetros del modelo **CASUAL** usando los datos obtenidos de Internet. Se usará para la tabla los datos generados en la descarga de *Facebook*, pues es una de las páginas más consultadas en la región según el ranking de alexa.com utilizando la fórmula de capacidad equivalente para N fuentes de tráfico. El



número de fuentes corresponde a los usuarios de Internet en cada país. Este procedimiento asume que un usuario descargando una página corresponde a una fuente de tráfico web y por tanto N usuarios corresponden a N fuentes.

Tabla 14. Capacidad requerida para el servicio web usando los parámetros de Facebook

Página	Parámetros y capacidad							Capacidad Kbps
	Usuarios	Servicio	Acceso Kbps	Tonline	Nc	Ns	Nr	
México	42.000.000,00	100%	2.512,00	25,8	1.505.000,00	10.595,62	5.736,92	14.411.144,07
Colombia	25.000.000,00	100%	2.176,00	26,6	923.611,11	6.537,08	3.564,64	7.756.664,60
Ecuador	3.352.000,00	100%	1.582,30	23,1	107.543,33	797,07	461,32	729.953,23
Perú	9.973.244,00	100%	1.888,00	27,4	379.537,34	2.723,02	1.511,73	2.854.137,25
Chile	10.000.000,00	100%	17.494,06	25,3	351.388,89	2.524,79	1.404,38	24.568.339,66
Bolivia	2.165.723,30	100%	1.167,92	23,1	69.483,62	523,58	309,33	361.274,79
Paraguay	1.500.000,00	100%	952,47	23,1	48.125,00	368,75	222,34	211.776,91
Uruguay	1.500.000,00	100%	1.064,13	23,1	48.125,00	368,75	222,34	236.604,17
Argentina	30.583.250,72	100%	2.200,00	28,7	1.219.082,35	8.601,03	4.670,30	10.274.662,45
Brasil	79.245.740,00	100%	2.048,00	27,1	2.982.721,60	20.882,14	11.221,72	22.982.092,80
Venezuela	10.976.342,00	100%	1.800,00	22,6	344.535,18	2.476,50	1.378,22	2.480.788,54

En la **tabla 14**, se muestra un estimado de la capacidad equivalente para N fuentes de tráfico usando para todos los países, los parámetros de descarga de un perfil de www.facebook.com, mostrado en la **tabla 11**, con el objetivo de observar cómo afecta el eje de usuarios en la demanda de tráfico web.

5.6.3. Capacidad equivalente del servicio web en distintos años de los países en estudio

En la **tabla 15**, se muestra el cálculo de la capacidad equivalente requerida para atender el servicio de web, para N usuarios de Internet en distintos años y para los distintos países en estudio.

En este caso la variable fundamental es el número de usuarios, que permite estimar el ancho de banda requerido para satisfacer las solicitudes de páginas web realizadas por esos usuarios. De hecho cualquier predicción sobre el número de usuarios de Internet



puede utilizarse en el modelo, proporcionando de esta manera la capacidad equivalente requerida para los usuarios predichos. El valor de capacidad que aparece en la tabla corresponde sólo al servicio web.

Tabla 15. Capacidad requerida para atender el servicio web en cada uno de los países

País	Unidad	Años				
		2007	2008	2009	2010	2011
México	Gbps	59,45	75,94	88,97	104,00	126,72
Colombia	Gbps	7,85	25,08	47,96	62,23	99,55
Ecuador	Gbps	1,59	2,37	3,01	5,73	8,96
Perú	Gbps	4,08	5,66	13,77	21,40	31,24
Chile	Gbps	32,73	50,38	73,20	174,77	290,06
Bolivia	Gbps	1,09	1,27	1,87	3,11	4,10
Paraguay	Gbps	0,53	0,69	1,23	1,63	2,34
Uruguay	Gbps	0,61	0,87	1,57	2,16	2,76
Argentina	Gbps	46,07	57,90	73,73	80,36	109,10
Brasil	Gbps	46,76	71,35	154,17	200,30	264,53
Venezuela	Gbps	6,60	12,23	19,52	22,54	31,05

5.6.4. Capacidad total a partir del servicio web

En la tabla 16, se muestra el cálculo de la capacidad total requerida para atender los distintos servicios de Internet. Para esto se parte del tráfico generado por el servicio web, que corresponde a un porcentaje del tráfico total que circula por Internet.

Con la información correspondiente al porcentaje de tráfico generado por el servicio web con respecto al tráfico total que circula por Internet, obtenida por medio de los distintos informes anuales de compañías como *Akami*, *Telegeography* y *Cisco*, puede estimarse para los distintos años, un aproximado del valor total de la capacidad mínima requerida para atender todos los servicios. Los cálculos mostrados en la tabla 16 se hicieron siguiendo la formulación mostrada en 5.6.1.1.

1 8 0 3



Tabla 16. Capacidad total demandada

País	Unidad	Años				
		2007	2008	2009	2010	2011
Servicio Web	%	46%	33%	28%	25%	17%
México	Gbps	129,25	230,11	317,75	416,01	745,39
Colombia	Gbps	17,07	75,99	171,30	248,92	585,62
Ecuador	Gbps	3,47	7,19	10,76	22,92	52,72
Perú	Gbps	8,86	17,15	49,17	85,60	183,74
Chile	Gbps	71,16	152,68	261,44	699,09	1.706,22
Bolivia	Gbps	2,37	3,86	6,67	12,46	24,14
Paraguay	Gbps	1,15	2,09	4,40	6,51	13,79
Uruguay	Gbps	1,34	2,62	5,61	8,64	16,22
Argentina	Gbps	100,15	175,45	263,32	321,43	641,75
Brasil	Gbps	101,64	216,20	550,61	801,22	1.556,04
Venezuela	Gbps	14,34	37,06	69,71	90,18	182,63

Un análisis interesante resulta de comparar el impacto sobre la capacidad, de la distribución de suscriptores por velocidades de acceso y la media ponderada de dicha distribución. Esta distribución se puede consultar en los distintos reportes de conectividad que hacen los ministerios de telecomunicaciones de cada país.

En la **tabla 16**, se presenta el cálculo de la capacidad de ancho de banda requerida por los usuarios de Internet en un país, este dato extrapolado a partir del servicio web. Es evidente que al tomar fielmente la distribución de usuarios por velocidad de acceso, el valor de la capacidad es menor, lo que quiere decir que la mayoría de los usuarios utiliza para acceder a la Red velocidades menores a la media ponderada, de ahí que el resultado con el acceso ponderado sea mayor. En los casos para los cuales la capacidad calculada con el acceso distribuido es mayor que la ponderada, se debe a que la desviación en el parámetros Nr en el modelo CASUAL, se hace muy grande debido a que las probabilidades de simultaneidad de la ráfagas son muy pequeñas.



La versatilidad del modelo **CASUAL** radica en que tiene en cuenta tanto el tráfico como la demanda por velocidades de acceso, de tal manera que no sólo es un modelo de tráfico, sino también un modelo de demanda.

Tabla 17. Capacidad requerida según el tipo de acceso en el modelo CASUAL

País	Modelo	Tipo de acceso	Años		Unidades
			2010	2011	
México	CASUAL	Distribuido ³⁸	330,01	559,70	Gbps
		Ponderado ³⁹	416,61	720,94	
Colombia	CASUAL	Distribuido	209,16	342,69	Gbps
		Ponderado	266,36	437,31	
Ecuador	CASUAL	Distribuido	22,73	50,96	Gbps
		Ponderado	22,82	52,50	
Perú	CASUAL	Distribuido	87,18	210,61	Gbps
		Ponderado	100,87	252,73	
Chile	CASUAL	Distribuido	976,75	2.100,86	Gbps
		Ponderado	1.236,23	2.691,38	
Bolivia	CASUAL	Distribuido	12,86	24,86	Gbps
		Ponderado	13,12	27,67	
Paraguay	CASUAL	Distribuido	5,74	15,18	Gbps
		Ponderado	5,11	14,49	
Uruguay	CASUAL	Distribuido	7,41	17,70	Gbps
		Ponderado	6,43	16,51	
Argentina	CASUAL	Distribuido	318,32	595,49	Gbps
		Ponderado	404,58	770,29	
Brasil	CASUAL	Distribuido	705,22	1.345,07	Gbps
		Ponderado	905,04	1.738,87	
Venezuela	CASUAL	Distribuido	75,02	149.017,82	Gbps
		Ponderado	92,81	182,68	

El modelo **FBM**, a diferencia del modelo **CASUAL**, es un modelo netamente de tráfico, y por tanto, no tiene en cuenta la demanda, ni tampoco las limitaciones de acceso. Este

³⁸ Tipo de acceso distribuido se refiere a que la estimación del ancho de banda requerido ha tenido en cuenta la distribución de los usuarios por velocidad de acceso

³⁹ Tipo de acceso ponderado se refiere a que la estimación del ancho de banda requerido se ha realizado con la media ponderada de la distribución de suscriptores por velocidad de acceso



modelo supone que el usuario tiene toda la disponibilidad de capacidad de red al momento de generar tráfico, en este caso la descarga de una página web. La demanda generada es función únicamente de la naturaleza de los datos, y dicha información está implícita en la traza empleada para alimentarlo; si la media de los datos es mayor entonces necesariamente el ancho de banda mínimo para atender al usuario se incrementa. Esta es la razón por la cual se alimentó usando trazas correspondientes al escenario en el que un usuario descarga una página, lo cual implica el mínimo uso posible de su conexión, y arroja cifras conservadoras. Si se pretendiera emplearlo en otro escenario habría que asegurarse de que a todos los usuarios considerados les sea posible efectuar el mismo proceso de descarga sin importar el ancho de banda que tengan disponible; es decir, se basa en suponer que todos los usuarios pueden generar la misma demanda de tráfico.

Los resultados consignados en la **tabla 18** en ningún momento son comparables, pues la naturaleza de ambos modelos es distinta, al igual que los datos con que se alimentan. Por tanto los valores obtenidos con el modelo **FBM** no sirven como referencia para calcular el error en el modelo **CASUAL** y lo mismo aplica en sentido contrario. Sin embargo se ha utilizado el modelo **FBM** por dos razones: la primera, tiene en cuenta los momentos activos e inactivos del tráfico, o en otras palabras los estados On-Off. La segunda, es que sus resultados están basados en el cálculo de la media, al igual que el modelo **CASUAL**. Estas dos similitudes, permiten relacionar los dos modelos, más no compararlos.

Para alimentar el modelo **FBM**, se han descargado 150 páginas con distinto tipo de contenido, la traza⁴⁰ de tráfico más representativa en cuanto a número de objetos, tamaño de descarga, latencia y tiempo de descarga, se obtuvo en la descarga de un perfil de *Facebook*, en este sentido la traza usada tenía características similares a los datos obtenidos para el modelo **CASUAL**. La traza obtenida representaba una fuente tráfico autosimilar. El algoritmo usado para el modelo **FBM** permitía calcular N fuentes con las mismas características. Para reemplazar el valor de N se usó el mismo parámetro

⁴⁰ Para obtener la traza se usó un programa de libre distribución llamado *Wireshark*,



de usuarios del lado del cliente utilizando en el nivel de conexión del modelo **CASUAL**. Los resultados pueden verse en la **tabla 17**.

Tabla 18. Capacidad requerida en FBM y CASUAL⁴¹

País	Modelo	Años			Unidades
		2009	2010	2011	
México	FBM	355,90	425,48	789,88	Gbps
	CASUAL	259,41	416,61	720,94	
Colombia	FBM	256,01	299,27	484,94	Gbps
	CASUAL	158,92	266,36	437,31	
Ecuador	FBM	27,03	42,10	76,25	Gbps
	CASUAL	3,22	22,82	52,50	
Perú	FBM	103,27	129,71	206,36	Gbps
	CASUAL	25,43	100,87	252,73	
Chile	FBM	106,60	130,97	206,86	Gbps
	CASUAL	556,86	1.236,23	2.691,38	
Bolivia	FBM	16,46	19,22	47,65	Gbps
	CASUAL	5,55	13,12	27,67	
Paraguay	FBM	14,25	19,26	37,11	Gbps
	CASUAL	0,87	5,11	14,49	
Uruguay	FBM	20,39	23,81	37,06	Gbps
	CASUAL	4,56	6,43	16,51	
Argentina	FBM	312,22	361,42	585,89	Gbps
	CASUAL	244,46	404,58	770,29	
Brasil	FBM	796,18	941,52	1.441,82	Gbps
	CASUAL	597,50	905,04	1.738,87	
Venezuela	FBM	112,18	131,66	225,45	Gbps
	CASUAL	65,64	92,81	182,68	

Aunque los datos obtenidos, no son comparables desde el punto de vista metodológico, en ellos puede observarse una tendencia en la demanda de tráfico, por ejemplo, si se mira esta tendencia desde en el modelo **FBM**, teniendo presente que su parámetro principal es la media de tráfico requerida, puede observarse en algunos casos que la

⁴¹ Es importante aclarar que esta no es una tabla comparativa de resultados. Simplemente se observa el comportamiento de ambos modelos desde los usuarios, pues este es un parámetro que tienen en común. Se ha optado por usar la media ponderada para resaltar solamente el impacto de los usuarios.



media de tráfico requerida está subestimada, esto sugiere que la capacidad requerida puede ser mayor, en otros caso está sobre estimada y por tanto se requiere menor capacidad. Se entiende entonces que los valores son cotas de la demanda de tráfico.

5.7. Adquisición de la base de datos

La adquisición de la base de datos para este modelo es una tarea delicada, pues son muchos datos y de muchas fuentes distintas. El proceso se detalla claramente dentro del manual que acompaña el aplicativo de implementación del modelo, al igual que la base de datos recopilada.

Esta base de datos se ha tomado de fuentes que se consideran fiables. El listado de sitios web más consultados por país, el tiempo que permanecen los usuarios dentro del sitio web, el número de páginas que visita dentro del sitio y la audiencia correspondiente a cada país se toma del sitio web www.alexa.com que es propiedad de la compañía *Amazon*.

Otros parámetros como tiempo total de descarga de la página, tamaño descargado, número de objetos descargados y latencia son consultados a través del sitio web www.pingdom.com que ofrece una herramienta online para la visualización de los datos.

Los usuarios de Internet y suscriptores de acceso, son tomados de distintos boletines oficiales y de empresas reconocidas, como la **CIA**, la **ITU**, *Internet World Stat* y los distintos ministerios encargados de las **TICs** en cada país.

Por último, el dato correspondiente a los porcentajes de tráfico que ocupa cada servicio de Internet, **VoIP**, **VoD**, **P2P**, Video, Web entre otros, se toman de los reportes semestrales y anuales que entregan compañías como **CISCO**, *Telegeography* y *Akamai*. Para el año 2011 se ha preferido usar el porcentaje presentado por *Akamai* debido la estimación de ese valor corresponde a mediciones hechas en sus servidores, ellos reportan que el 17% de las solicitudes hechas a sus servidores es de tráfico web.

Se ha preferido el dato de *Akamai* sobre los de **CISCO** y *Telegeography*, porque no es una medición sobre nodos de red y por tanto se evita el sesgo generado por la



ubicación geográfica de los nodos y además es más cercano al tráfico de fuentes web generado por los usuarios.



CONCLUSIONES

De los modelos revisados en este trabajo, el modelo **CASUAL** resulta ser muy versátil porque permite analizar los distintos factores que componen el tráfico que circula por Internet. Por un lado permite capturar la demanda de los usuarios de Internet a través del eje de usuarios, esto es muy importante porque son los usuarios los que generan el tráfico cuando interactúan con los distintos servicios y aplicaciones que ofrece la Red.

Otras de las ventajas que ofrece el modelo, es permite tratar cada eje por separado y ajustarlo a las necesidades particulares de un análisis, por ejemplo, si se quisiera hacer predicción de la capacidad equivalente debida solamente al crecimiento de los usuarios, se podría utilizar cualquier modelo predictivo sobre este eje, sin que esto afecte los demás. En este eje se podría usar el modelo de Bass para la predicción del crecimiento de los usuarios.

Por el lado del eje de accesos, permite limitar el tráfico dependiendo del ancho de banda que tenga el usuario para ingresar a Internet. Se puede hacer respetando el factor de multiplicidad Usuarios/suscriptores y la distribución de suscriptores por velocidad de acceso, o si se desea, una media ponderada. En este eje se podría usar la Matriz de Transición modulada por Markov, lo que permitiría proyectar la distribución de suscriptores por velocidades de acceso en distintos años.

Este un modelo mixto, que combina distintas distribuciones de probabilidad, permitiendo de esta forma estudiar la naturaleza de los datos sin tener en cuenta el tipo de servicio de Internet, sin embargo sus resultados mejoran cuando se analiza sólo las trazas de tráfico generadas por el servicio Web. Presenta una ventaja, y es, que permite hacer el análisis del tráfico, no sólo del lado del servidor, si no del lado del cliente sin perder la parsimonia.

Al usar un proceso de llegada poissoniano, se facilita el cálculo de las probabilidades *On-Off* de los distintos niveles, sin embargo esto hace pensar que la duración de las



sesiones y ráfagas es invariante, lo que en realidad no es cierto. Este hecho se convirtió en una ventaja fundamental a la hora de reinterpretar los parámetros, porque permitió utilizar valores de tráfico indirectos, obtenidos con aplicaciones libres, sin la necesidad de recurrir a mediciones de tráfico con equipos especializados de alto costo, cumpliendo de esta forma el objetivo de usar el modelo sin la necesidad de mediciones directas, y facilitar su uso para labores administrativas y bajo costo computacional.

El valor de capacidad de ancho de banda obtenido con este modelo, depende de la fidelidad de los datos usados para alimentar los parámetros. No hay forma de juzgar el error en la estimación de la capacidad, pues no se encontró un estudio, donde las distintas compañías **ISPs** publicaran el valor total de ancho de banda que tienen instalado en un país, por tanto este modelo sólo presenta una estimación genérica del ancho de banda demandado por los usuarios sujeta a unas condiciones de acceso y servicios. En este sentido resulta útil la **tabla 18** donde los resultados del modelo **FBM**, hacen pensar que las modificaciones hechas en el cálculos de las probabilidades de los distintos niveles del modelo **CASUAL**, son muy acertadas.



TRABAJOS FUTUROS

En el desarrollo de este trabajo, se encontraron temas de interés que han motivado la inclusión de esta sección. Estos temas no sólo desembocan en un conocimiento profundo de las características principales del tráfico, los servicios y las redes en general, sino en un mejor aprovechamiento de los recursos disponibles.

Un estudio importante, es incorporar al modelo casual, modelos de predicción, no sólo con el objetivo de ver como se incrementa a corto y largo plazo la demanda de tráfico debida a los servicios y aplicaciones para de esta manera adaptar la Red, sino también cómo pueden ayudar en distintas tareas de planificación estratégica.

Análisis de costos de **GoS** (*Grade of Service*) y **QoS** en redes multiservicio, donde los costos incrementales a largo plazo sólo consideran el ancho de banda medio y no el impacto de tráfico de los distintos servicios sobre la red.

Análisis para la asignación óptima de los recursos en una red, por medio de procesamiento en paralelo y redes con direccionamiento inteligente.

Análisis del impacto de la penetración de nuevas tecnologías en la predicción de la demanda de tráfico y velocidades de acceso.

Estudiar como diagramas de estados, redes neuronales, árboles de decisión, algoritmos genéticos pueden usarse como herramientas para la estimación de valores óptimos para **ToS** (*Terms of Service*) y control de **QoS** adaptativo que puede servir por ejemplo para prever la congestión de un par de nodos de la red y redireccionar automáticamente los paquetes.

UNIVERSIDAD
DE ANTIOQUIA
1803



BIBLIOGRAFIA

1. **GOGL, H. GREINER, M. Y SCHWEFEL, H. P.** *Model Calibration*. München : Rechnerstruktur/-architektur, 1997.
2. **Comisión de estudio 2/16 del UIT-D y Congresos internacionales de teletráfico, UIT.** *Manual sobre ingeniería de teletráfico*. Ginebra : UIT, 2002.
3. **Desconocido.** *Técnicas de modelado de tráfico y calidad de experiencia en servicios multimedia*. s.l. : UIT, 2010.
4. **Monk, T. y Claffy, k.** *A survey of Internet statistics / metrics activities*. San Diego : National Laboratory for Applied Network Research (NLNAR), 1996.
5. **REYES LECUONA, Arcadio.** *Modelado de Tráfico de Clientes WWW*. Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad de Málaga. Málaga : Universidad de Málaga, 2001. Tesis doctoral.
6. *An architecture for large scale Internet measurement*. **Paxson, V., Mahdavi, J., Adams, A. y Mathis, M.** s.l. : IEEE Communications, Agosto de 1998, IEEE Communications Magazine, Vol. 36, págs. 48-54.
7. **CAIDA.** CAIDA. *ISMA - Internet Statistics and Metrics Analysis Workshops*. [En línea] University of California's San Diego Supercomputer Center, 1996. <http://www.caida.org/workshops/isma/>.
8. **García Gutiérrez, Alberto Eloy.** *Contribución al desarrollo de herramientas estratégicas para el diseño, dimensionado y evaluación de redes de telecomunicación de banda ancha*. Departamento de Ingeniería de Comunicaciones , Universidad de Cantabria. Santander : s.n., 2009. pág. 279, Tesis doctoral.
9. **Vidal, Jorge García.** *Modelos analíticos para la evaluación de mecanismos de control de tráfico en redes ATM*. Departamento de Arquitectura de Computadores, Universidad Politécnica de Catalunya. Barcelona : s.n., 1992. Tesis doctoral.
10. **Pallarés Segarra, Esteve.** *Contribución a la evaluación y dimensionado de nodos y enlaces en redes de alta velocidad*. Escuela Técnica de Ingeniería de Telecomunicaciones de Barcelona, Universidad Politécnica de Catalunya. Barcelona : s.n., 2001. Tesis doctoral.
11. *Traffic modeling for telecommunications networks*. **Frost, V., y Melamed, B.** Lawrence, KS, USA : IEEE, Marzo de 1994, Communications Magazine, IEEE, Vol. 32, págs. 70-81.

12. *Introducción a los modelos de tráfico para redes de Banda Ancha*. **Rincón R., David**. 11, Barcelona : IEEE BCN Student Branch, Mayo de 1998, Buran, págs. 41-48.
13. *Modelos de tráfico en análisis y control de redes de comunicaciones*. **Alzate, Marco A.** 1, Bogotá : Universidad de los Andes, 2004, Ciencia, Investigación, Ciencia y Desarrollo, Vol. 9, págs. 63-87.
14. **Belzarena, Pablo**. *Ingeniería de tráfico en línea en redes MPLS aplicando la teoría de grandes desviaciones*. Facultad de Ingeniería, Universidad de la República Uruguay. Montevideo : s.n., 2003. Tesis de maestría.
15. *Series E: Overall Network Operation, Telephone Service, Service Operation and Human Factor*. **ITU, International Telecommunications Union**. Recomendación E-490.1, s.l. : ITU-T Study Group 2, Enero de 2003.
16. *Tráfico Autosemejante*. **Pérez Mejía , Juan Andrés y Romero Parra, Jorge Mauricio**. 3, Bogotá : Universidad Santo Tomás, 2005, Hallazgos - Revista de Investigaciones, págs. pp. 119 - 134.
17. **Mordecki, Ernesto**. *Distribución del máximo de un proceso de Lévy y aplicaciones en finanzas y matemática actuarial*. Facultad de Ciencias, Centro de Matemáticas. Montevideo, Uruguay : s.n., 2003. Notas de Curso. Curso de análisis estocástico y aplicaciones, 28 de julio. Disponible en: www.cmat.edu.uy/~mordecki.
18. **Rodríguez Hernandez, Juan Gabriel**. *Una introducción a los modelos de series temporales no lineales*. Universidad Rey Juan Carlos. Madrid : s.n., 2001. Pág. 34.
19. *Tráfico vehicular en zonas urbanas*. **Lozano, Angélica, Torres , Vicente y Antún, Juan Pablo**. 70, Distrito Federal : Universidad Nacional Autónoma de México, Abril - Junio de 2003, Ciencias, págs. 34 - 45.
20. **Aracena Urrutia, Carlos y Veliz Portuguez, Alejandro**. *Desarrollo del modelo trace sampling para generar tráfico en redes IP*. Departamento de Ingeniería Eléctrica, Universidad de Santiago de Chile.
21. *Introducción al tráfico autosimilar en redes de comunicaciones*. **Alzate Monroy, Marco Aurelio**. Nº 2, s.l. : Grupo Editorial Gaia, 2001, Colombia Ingeniería, Vol. Vol. 6, págs. pp. 6 -16 .
22. *Aplicaciones del modelo On/Off al tráfico agregado en las redes de comunicaciones*. **Parra León , Andrés, Piedrahita, Elkin M y Salcedo, Octavio**. 30, Bogotá : Universidad Distrital Francisco José de Caldas, 14 de diciembre de 2011, Tecnura, Vol. 15, págs. pp. 129 - 147.
23. **Fitzgerald, Jerry y Dennis, Alan**. *Redes y comunicación de datos en los negocios*. 3ª ed. México : Limusa Wiley, 2003. Pág. 472 p. ISBN: 968-18-6351-8.
24. **Luján Mora, Sergio**. *Programación de aplicaciones web, historia , principios básicos y clientes web*. Alicante : Editorial club universitario, 2009. Pág. 354 p. ISBN: 84-8454-206-8.



25. **St-Pierre, Armand y Stéphanos, William.** *Redes locales e Internet, introducción a la comunicación de datos.* reimp. 2003. México : Trillas, 1997. pág. 378. ISBN: 968-24-0904-7.
26. **Orfali, Robert, Harkey , Dan y Edward, Jeri.** *Cliente/servidor y objetos, guía de supervivencia.* 3ª ed. México : Oxford University Press México S.A., 2002. pág. 850. ISBN: 970-613-597-9..
27. **Forouzan, Behrouz A.** *Transmisión de datos y redes de comunicaciones.* Madrid : McGraw-Hill/Interamericana de España S.A.U., 2007. pág. 902. ISBN: 978-84-481-5617-6.
28. **Sheldom, Tom.** *Enciclopedia LAN Times de redes.* Madrid : McGraw-Hill/Interamericana de España S.A.U., 1994. pág. 1188. ISBN: 84-481-1787-5.
29. **Halsall, Fred.** *Redes de computadores e Internet.* 5ª ed. Madrid : PEARSON EDUCACIÓN, 2006. pág. 856. ISBN: 978-84-7829-083-3.
30. **León-García, Alberto y Widjaja, Indra.** *Redes de comunicación, conceptos fundamentales y arquitecturas básicas.* Madrid : McGraw-Hill/Interamericana de España S.A., 2002.
31. **Tanenbaun, Andrew S.** *Redes de computadoras.* 4ª ed. Distrito Federal : Pearson Educación de México, 2003. pág. 892. ISBN: 970-26-0162-2.
32. **Stallings , William.** *Comunicaciones y redes de computadores.* 5ª ed. Madrid : PRENTICE HALL IBERIA, 1997. pág. 808. ISBN: 84-89660-01-8.
33. **Stallings, William.** *Redes de Internet de alta velocidad, rendimiento y calidad de servicio.* 2ª ed. Madrid : PEARSON EDUCACIÓN S.A., 2004. pág. 752. ISBN: 84-205-3921-X.
34. **Marino Pezzarini, Silvana María.** *Aprendizaje electrónico y auto-evaluación de TCP/IP.* Facultad de Ciencias Exactas, Naturales y Agrimensura, Universidad Nacional del Noreste. Corrientes, Argentina : s.n., 2007. pág. 256, Tesis de grado.
35. **Martinez Ruíz, Ignacio.** *Contribución a los modelos de tráfico y control de QoS en los nuevos servicios sanitarios basados en telemedicina.* Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza. Zaragoza : Instituto de Investigación en Ingeniería de Aragón, 2006. pág. 200, Tesis doctoral.
36. **Carballar, José.** *ADSL Guía del usuario.* Distrito federal : Alfa Omega Grupo Editor S.A., 2003.
37. **MinTIC, Ministerio de Tecnologías de la Información y las .** *Boletín trimestral de las TIC conectividad, cifras segundo trimestre de 2011.* Cundinamarca. Bogotá : s.n., 2011.
38. *Informe trimestral de conectividad.* Comisión de Regulación de Telecomunicaciones, República de Colombia. Bogotá : s.n., 2009. 14.



39. *Simplified methods for next generation IP Access network planning*. **García, Alberto Eloy y Hackbarth, Klaus D.** Madrid : Proceedings of the 5th WSEAS, 2006. Conf. on Electronics, Hardware, Wireless an Optical communications. págs. 100 - 107. February 15 - 17.
40. *Approximation to a behavioral model for estimating traffic aggregation scenarios*. **García, Alberto Eloy y Hackbarth, Klaus D.** Nº 5, Cantabria : s.n., 2008, Journal of Universal Computer Science, Vol. Vol. 14, págs. 731 - 744.
41. *Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level*. **Murray Hill, N.J., y otros, y otros.** s.l. : Journals & Magazines, February de 1997, Networking, IEEE/ACM Transactions on, Vol. Vol. 5, págs. 71 - 86.
42. **Crovella, Mark E.** Heavy-Tailed Probability Distributions in the World Wide Web. *In A Practical Guide To Heavy Tails*. s.l. : Chapman and Hall, 1998, Chapter 1, págs. 3-26.
43. **Bravo García, Patricia.** *Modelado fractal de tráfico de la red de Ingeniería Telemática*. Departamento de Ingeniería Telemática, Universidad Carlos III de Madrid. Madrid : Escuela Politécnica Superior, 2004. pág. 169, Tesis de grado.
44. **Ferreira , Eva y Garín, María Araceli.** *Estadística actuarial: modelos estocásticos*. Facultad de Ciencias Económicas y Empresariales, Universidad del País Vasco. Bilbao : Sarriko-On, 2010. pág. 75. Disponible en: <http://ekasi.edu.es>.
45. **Blaiotta, Jimena y Delieutraz, Pablo.** *Teorema central del límite*. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Buenos Aires : s.n., 2004. pág. 17.
46. **Rovira Escofet, Carles.** *Teorema central del límite*. s.l. : UOC. pág. 22, Notas de clase. En: www.uoc.edu.
47. **Casilari, E, y otros, y otros.** *Modelado de tráfico telemático*. Departamento de Tecnología Electrónica, Universidad de Málaga. Málaga : s.n. pág. 10, Artículo.
48. **Akamai.** *The state of the Internet*. 2011. pág. 48, Reporte. Vol. 4; Nº 3.

UNIVERSIDAD
DE ANTIOQUIA
1803

